UNIVERSITY OF LEEDS

Threonine Aldolases as Tools for Stereoselective Synthesis

by

Alex George Moloney

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
Astbury Centre for Structural Molecular Biology

September 2018

Declaration

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

The right of Alex George Moloney to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Alex George Moloney
September 2018

Acknowledgements

I would firstly like to thank my supervisors, Professor Alan Berry and Professor Adam Nelson. They have been extremely supportive during my four years as a PhD student. Special thanks must also go to Dr Marc Van der Kamp for his guidance, patience and expertise that have been fundamental to my final results chapter. Thanks to Dr Chi Trinh for his help with crystallography work and to Berry, Hemsworth and Radford Lab members (past and present) for their friendships and useful discussions. Thank you to my industrial supervisor, Dr Matthew Bycroft. I have enjoyed my visits down to Cambridge and the regular meetings keeping you up to date with my work.

Finally, I couldn't have finished this thesis without the love and support (and snacks) from my amazing partner Isobel. She has put up with my tantrums and strops and kept me on the straight and narrow during stressful times. Thank you to my Mum and Dad for believing in me and for and their undivided support. To you three, thank you for being my number one fans, I promise to do my best to remain fan-worthy!

Abstract

Enzymes are biological catalysts capable of an enormous array of reactions. They demonstrate the ability to be reshaped and repurposed as synthetic tools by the process of enzyme engineering. Industrially, enzymes have been exploited for their ability to synthesise high value chiral molecules with excellent stereoselectivity. They are capable of functioning at low temperatures, near neutral pH and are completely biodegradable. Methods to harness the synthetic potential of enzymes have been extensively developed and applied over recent years. Directed evolution is one such approach that mimics natures evolution strategy but over a significantly shorter time frame. It involves iterative cycles of mutagenesis and screening until a required function is achieved. Advances in structural biology have enabled a huge number of enzyme structures to become available. These can help to implicate important catalytic residues in an enzyme, in turn, focussing engineering efforts to likely mutagenic hotspots. This can lead to greater chances of obtaining a desirable enzyme function if paired with an effective mutagenesis and screening strategy.

In this thesis we use combinatorial active site saturation testing, a focused directed evolution method, to engineer a *low-specificity* L-threonine aldolase from *Escherichia coli* for improved stereoselectivity. We target the reversible retro-aldol cleavage reaction of phenylserine, an industrially interesting compound with four possible stereoisomers. Stereoselective variants are identified following the development of an effective high-throughput screen. Further, we use this screen to identify and produce rational combinations of variants. This achieves an enzyme that is 500-fold stereoselective for a single phenylserine stereoisomer. We also aim to provide insight into the poor stereoselectivity of the wild-type enzyme using molecular mechanics and quantum mechanics/molecular mechanics simulations. These are performed on an obtained 1.6 Å crystal structure of *Escherichia coli* threonine aldolase. We further propose plausible mechanisms for the stereogenic step of the aldol-condensation reaction.

Contents

List of Figures

List of Tables

Abbreviations

Å	Angstrom
A. jandaei	*Aeromonas jandaei*
C-C	carbon-carbon bond
d(C-C)	reaction coordinate for aldol carbon-carbon bond
CASTing	combinatorial active site saturation testing
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
D-TA	D-threonine aldolase
E. coli	*Escherichia coli*
epPCR	error-prone polymerase chain reaction
eTA	*E.coli* threonine aldolase
IPTG	isopropyl-β-D-thiogalactoside
ISM	iterative saturation mutagenesis
k_{cat}	catalytic constant
K_m	Michaelis constant
L-TA	L-threonine aldolase
MM	molecular mechanics
MD	molecular dynamics
PCR	polymerase chain reaction
PDB	Protein Data Bank
PLP	pyridoxal 5'-phosphate
PS	phenylserine
psec	picosecond
PXG	PLP-glycine ligand
QM	quantum mechanics
SDS PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
TA	threonine aldolase
TS	transition state
UV	ultraviolet
WHAM	weighted histogram analysis

Chapter 1

Introduction

1.1 Biocatalysis

Enzymes are macromolecular structures essential to the function of all living organisms. Fundamental to life, enzymes are biological catalysts (biocatalysts) with the ability to facilitate an extraordinary number of diverse reactions. Since the turn of the millennium, applications of biocatalyts have started to become more apparent. They offer a safe, cheap and green route for the synthesis of valuable molecules; features largely absent in their chemocatalyst counterparts. Furthermore, biocatalysts are capable of functioning with high chemo-, regio-, diastereo-, and enantioselectivity, near neutral pH and at low temperatures [Bornscheuer and Kazlauskas, 2006, Chen et al., 2017, Reetz, 2004]. By nature enzymes are evolvable meaning they can be reshaped to perform a specific function. Harnessing this evolvability by laboratory based approaches is broadly referred to as protein engineering and is considered the driving force for useful applications of biocatalysts.

As a process, biocatalysis describes these useful applications of enzymes and is said to be in its third wave [Bornscheuer et al., 2012, Turner and Truppo, 2013]. This grading is very much based on the rate at which synthetically useful biocatalysts can be produced. The first wave became apparent over 100 years ago as scientists used natural biological extracts to mediate a desired reaction. An example is the synthesis of *R*-mandelonitrile from benzaldehyde and cyanide with isolated plant extracts [Rosenthaler, 1908]. The second wave emerged in the late 1900's as the substrate scope of enzymes was realised. This was facilitated by improvements in structural biology guiding early protein engineering

efforts [Estell et al., 1985]. In this wave, biocatalysts were apparent in the resolution of chiral intermediates [Matsumae et al., 1993], the synthesis of agrochemicals [Griengl et al., 2000] and in the production of cholesterol-lowering statins [Müller, 2005]. Now well into the third wave, and possibly even approaching the fourth [Truppo, 2017], biocatalysis is imperative to the fine chemical, food and pharmaceutical industries. With respect to the pharmaceutical industry, biocatalysis has only recently cemented its place as a viable process. Typically, the lead times required for their development is considered too long and not worth the investment of time and money. However, this new era is equipped with advances in directed evolution methods and computation that are capable of accelerating the process (Section 1.2). Improvements in these directed evolution techniques are considered to be one of two factors associated with increasing biocatalyst implementation. The other is the ease of access to enzymes.

Almost one billion gene sequences from the genomes of 38,000 different organisms are now publicly available in the GenBank database[1]. As shown in Figure 1.1, the number of sequences continues to grow at an almost exponential rate, thanks to advances in sequencing technology. However, all this information provides is an incomprehensibly large number of sequences composed of different combinations of four nucleotide bases. We therefore require an approach to make use of such a database. Bioinformatics is a field focused on interpreting biological data *in silico*, using mathematical and statistical techniques [Diniz and Canduri, 2017]. Tools within this field can be utilised to mine databases, returning useful information. Novel and potent biocatalysts can be identified through structural alignments of coding sequences of enzymes with known function [Höhne et al., 2010]. Key information relating to conserved amino acid positions in ancestral families of enzymes can be obtained. In turn this has lead to biocatalysts with broader substrate range, promiscuous activities, enhanced stabilities and altered stereoselectivities [Diniz and Canduri, 2017, Fesko et al., 2018, Kamerlin and Warshel, 2010, Jochens and Bornscheuer, 2010].

The repetoire of biocatalytic reactions extends to (but is not limited to): reductions, oxidations, transaminations and condensations. All of these have synthetically useful applications (Figure 1.2) [Urano et al., 2011, Chen et al., 2010, Jennewein et al.,

[1] Statistics from https://www.ncbi.nlm.nih.gov/genbank/statistics/

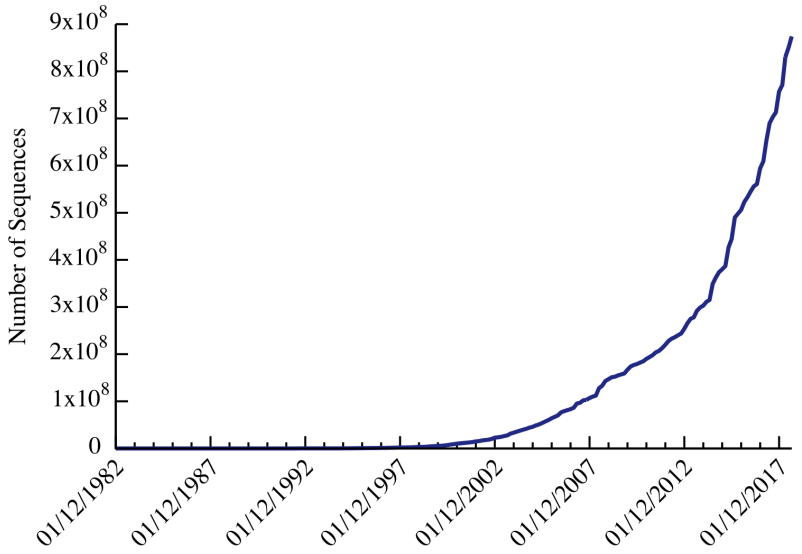


Figure 1.1: Number of gene sequences available in GenBank since its release in 1982.

2006]. Urano et al. [Urano et al., 2011] successfully engineered an amino alcohol dehydrogenase for optimal production of D-pseudoephadrine (used in decongestant and anti-asthmatic compounds). Their enzyme catalysed the stereoselective reduction of (*S*)-1-phenyl-1-keto-2-methylaminopropane into D-pseudoephedrine with three fold greater efficiency than the wild-type (Figure 1.2A).

In another example, Chen et al. [Chen et al., 2010] developed an efficient enzymatic route to an anti-diabetic drug candidate ((*S*)-2-amino-3-(6-*o*-tolylpyridin-3-yl)propanoic acid) (Figure 1.2B). Using a one-pot, two-enzyme system they achieved a 66% isolated yield of (*S*)-2-amino-3-(6-*o*-tolylpyridin-3-yl)propanoic acid and 99.9% e.e. under industrially relevant conditions. The two-step reaction first involved an *R*-selective oxidation of the racemic amino acid substrate, forming the alpha keto acid with a subsequent transamination to the target compound.

The synthetic use of an aldolase enzyme was demonstrated for 2-deoxy-D-ribose 5-phosphate aldolase (DERA) [Jennewein et al., 2006]. Through directed evolution a 10-fold improved stereoselective DERA variant was obtained, with respect to the wild-type enzyme, for the aldol-condensation of 2-deoxy-D-ribose 5-phosphate from acetaldehyde and D-glyceraldehyde 3-phosphate (Figure 1.2C). The target compound is used as a key chiral intermediate in the synthesis of vastatin drugs [Jennewein et al., 2006].

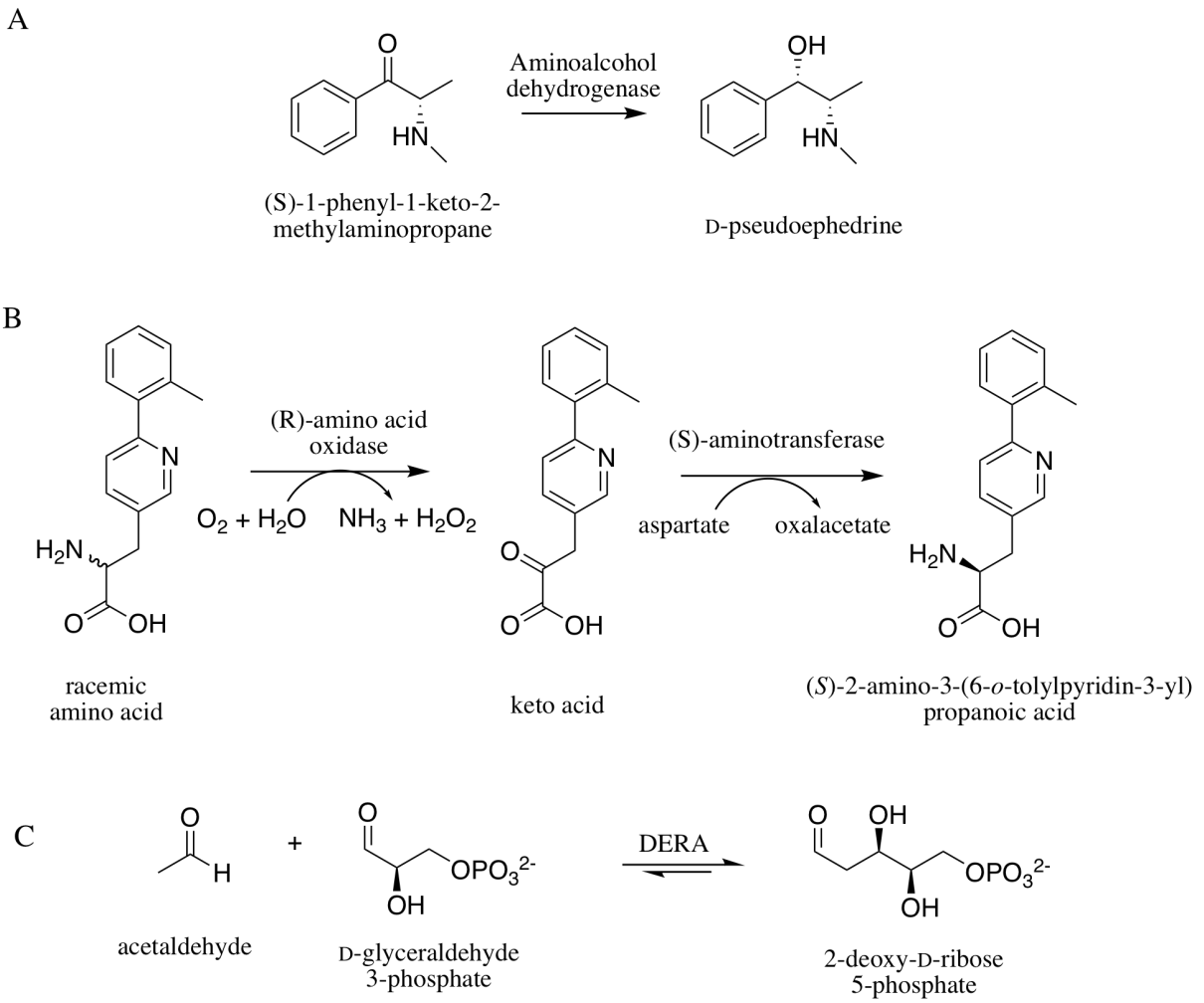


Figure 1.2: Industrially implemented biocatalyts. A: The stereoselective reduction by an aminoalcohol dehydrogenase [Urano et al., 2011]. B: Stereoselective oxidation by an *R*-aminoacid oxidase and a stereoselective transamination by a *S*-aminotransferase [Urano et al., 2011]. C: The stereoselective aldol-condensation of an engineered 2-deoxy-D-ribose 5-phosphate aldolase (DERA) [Jennewein et al., 2006].

One excellent example of a biocatalyst from the third wave [Bornscheuer et al., 2012] is the engineered transaminase for the synthesis of sitagliptin [Savile et al., 2010]. Used as a dipeptidyl peptidase-4 inhibitor in the treatment of type II diabetes [Williams-Herman et al., 2008], sitagliptin required a complex chemical manufacturing scheme. This involved high pressure and a rhodium-based chiral catalyst [Hansen et al., 2009] (Figure 1.3A). The requirement of such a catalyst added significant cost to the process and furthermore reduced yield due to necessary additional purification steps. Savile et al. [Savile et al., 2010] used a combination of *in silico* design and directed evolution to achieve a highly selective transaminase (Figure 1.3B). The problematic chemocatalytic steps were subsequently bypassed with a safer, greener, cheaper and higher yielding biocatalytic route.

The engineered enzyme provided sitagliptin with a 10-13% increase in overall yield, a 53% increase in productivity (kg L^{-1} per day), a 19% reduction in total waste, elimination of heavy metals and a reduction in total manufacturing cost [Savile et al., 2010]. Remarkably, the final implemented transaminase contained 27 amino acid substitutions with respect to the wild-type enzyme. Four of these were identified from structure guided *in silico* docking, eight from saturation mutagenesis, five from random mutagenesis and ten from homology libraries. Even more remarkably this variant was achieved over the space of a single year. With this example it is easy to see the need for industrial biocatalysts. Given the available tools in the current wave, we might expect an increasing number of chemocatalyts to be replaced by their biological counterparts.

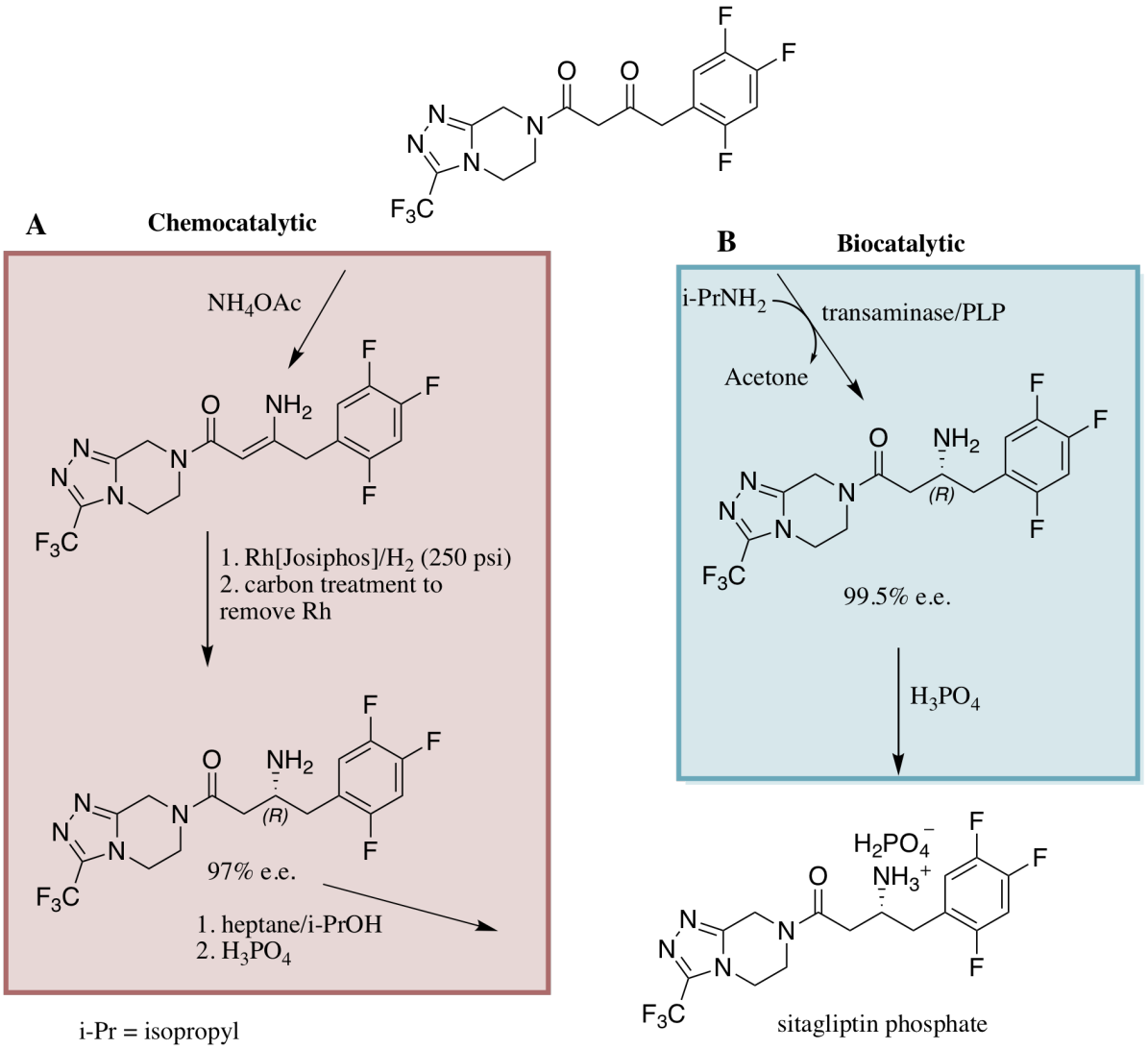


Figure 1.3: Chemocatalytic and biocatalytic routes to sitagliptin phosphate. A: The chemocatalytic route involves enamine formation followed by asymmetric hydrogenation at high pressure (250 psi) with a rhodium catalyst. This yields sitagliptin at 97% e.e. with trace amounts of rhodium. B: The biocatalytic route implements the engineered transiminase for direct amination of the presitagliptin keytone, providing enantiopure sitagliptin (99.95% e.e.). Figure adapted from [Savile et al., 2010].

1.2 Directed Evolution

Since the origin of life, organisms have been challenged by selection pressures forcing them to adapt and change. We refer to these adaptations as evolution. Today, we see diverse classifications of living organisms due to many iterations of evolution. The relationship between genotype and phenotype is imperative where changes in the DNA prove desirable when expressed in an organism. For centuries humans have used methods such as cross-breeding to select and enhance the quality of livestock and crops. Modern day

advances in genetics and molecular biology allow for targeted, single-gene evolution. As a process, directed evolution can be separated into two parts. The first involves generating genetic diversity in a gene of interest using a mutagenesis strategy. Secondly, variants must be screened with an effective screen linking genotype to phenotype. Once variants with desirable traits are identified the process is repeated as many times as necessary (Figure 1.4).

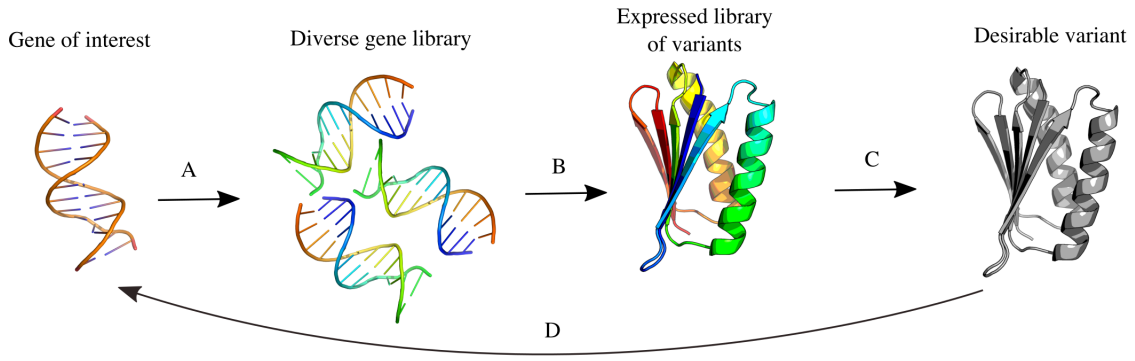


Figure 1.4: A generalised schematic of directed evolution. A gene of interest for an enzyme of particular function is first selected. A: A library of genetic variants is created using random or focused methods. B: Genes are expressed to produce enzyme variants. C: Desirable or improved variants are identified by a screen linking genotype and phenotype. D: Iterations of the process are performed for further improvements in function as many times as necessary.

Many methods for generating genetic diversity exist in protein engineering. These are considered to be either, random, homology guided or focussed based approaches and shall be discussed in the following sections.

1.2.1 Protein Engineering

Random mutagenesis is considered the most effective method to apply in the absence of structural information. One of the most well used methods is error-prone PCR (epPCR) [Leung, 1989, Chen and Arnold, 1993]. It was discovered that under certain conditions, a low-fidelity DNA polymerase would introduce random point mutations across a gene. The identified conditions involved variations in magnesium concentration, supplementation with manganese and use of mutagenic dNTP analogues. It was found that error rates of between 10^3-10^4 could be achieved by this method [Wilson and Keefe, 2000]. As mutations are introduced during the PCR elongation step, it is possible to tune the mutational

frequency by the number of cycles [Wilson and Keefe, 2000] (Figure 1.5). The method first demonstrated success in 1993 when Chen and Arnold evolved the protease, subtilisin E, to function in a highly non-natural environment [Chen and Arnold, 1993]. They performed three successive rounds of random mutagenesis to identify a variant that functioned 256-times more efficiently than the wild-type enzyme in 60% dimethylformamide.

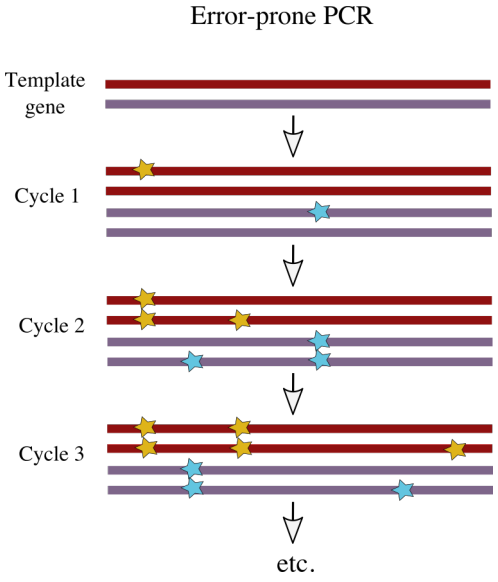


Figure 1.5: A schematic representation of error-prone PCR. Random point mutations accumulate through successive cycles of elongation. The mutational frequency can be tuned by the conditions of the reaction and the number of cycles.

Around the same time as the emergence of epPCR an alternative, recombinatory method for generating gene diversity was identified. Stemmer et al. [Stemmer, 1994a] described the DNA shuffling method where families of homologous genes are fragmented and recombined by PCR to create large mutagenic libraries. While not strictly random, diverse libraries are generated through homology guidance. Stemmer et al. [Stemmer, 1994b] demonstrated the effectiveness of this method with a TEM-1 β-lactamase. Three iterations of DNA shuffling were performed with subsequent screening. This identified a variant with a 32,000-fold increase in the minimum required inhibitory concentration for the antibiotic cefotaximine [Stemmer, 1994b]. The variant was 2000-fold more efficient than an epPCR approach targeting the same system [Palzkill and Botstein, 1992]. This approach has since been widely used in directed evolution. A further example is observed for an engineered aldolase with altered stereoselectivity [Williams et al., 2003]. Similarly, through three rounds of DNA shuffling and screening a N-acetylneuraminic acid lyase variant with an

80-fold improvement in k_{cat}/K_m for a non-natural, fructose 1,6-bisphosphate substrate was identified. This manifested as a 100-fold change in the stereoselectivity of the reaction.

While epPCR and recombinatory methods significantly differ in generating library diversity, they can be combined for great effect. In one example an engineered subtilisin E variant was obtained with a 50-fold improvement in thermostability (measured as increased half-life at 65°C) [Zhao et al., 1998]. The two part engineering effort first obtained five thermostable variants by epPCR. Secondly, these were recombined using a staggered extension process (stEP) (Figure 1.6). This involved a modified PCR protocol where the elongation step is extremely abbreviated with heat denaturation. As the temperature is lowered the short elongated fragments anneal to complementary full length templates for further elongation. Slowly these fragments increase in size until full length genes are obtained.

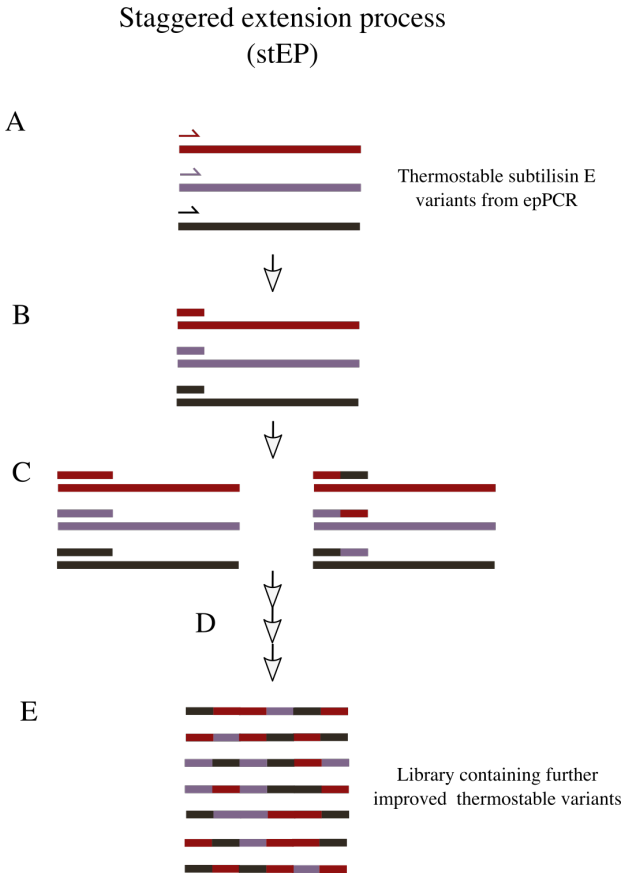


Figure 1.6: The staggered extension process (StEP) recombination for an improved theromostable subtilisin E using a modified PCR protocol. Single primers and single strands of DNA for the epPCR variants are shown. A: Denatured template genes are primed with one defined primer. B: Short fragments are produced in an abbreviated elongation. C: A further cycle of stEP allows fragments to randomly prime to a template and extend further. D: StEP cycles are repeated until full-length genes are produced. E: A library of full length gene variants is obtained for subsequent screening. Figure adapted from [Zhao et al., 1998].

1.2.2 Focused mutagenesis

An enzyme 100 residues in length has $1.26x10^{130}$ possible combinants. It would be impossible to screen even a single percentage of these. However, this "numbers problem" is a well addressed topic in enzyme engineering. Previous successes in directed evolution would suggest engineering efforts be focused around the active site of an enzyme [Reetz et al., 2001]. Given there are close to 150,000 structures available in the PDB[2] we might expect such information to be available for a particular enzyme or a homologue. In turn, this can be used to implicate key residues in enzyme catalysis, narrowing engineering efforts. However, saturation of just eight residues still yields $2.6x10^{10}$ possible combinations; still an apparent "numbers problem". To apply some perspective, that would require screening of around eight mutants a second, non-stop, for 100 years to fully sample the entire sequence space.

1.2.2.1 Combinatorial active site saturation testing (CASTing)

Combinatorial active site saturation testing (CASTing) is a method developed by Reetz et al. [Reetz et al., 2005, 2006] targeting this "numbers problem". It uses knowledge-driven systematisation of focused libraries around the binding pocket. The method couples an effective mutagenesis strategy with an efficient screen to enable sampling of a huge, focused, sequence space. Libraries are generated through the use of degenerate codons. These are codons constructed with a mixed population of nucleotides at a given position. As well as the four standard DNA bases (A, T, C and G), several other letters are used with combinations of each standard base assigned (Table 1.1).

By the CASTing approach, multiple residues are grouped and targeted for saturation mutagenesis with degenerate codons. This enables synergistic combinations between multiple residues to be identified. When a variant with desirable function is obtained, iterative rounds of CASTing are performed with this as the new parent template. This process is termed iterative saturation mutagenesis (ISM) [Reetz et al., 2006]. A hypothetical CASTing and ISM strategy is shown in Figure 1.7.

[2] Statistics from https://www.rcsb.org/stats/growth/overall

Degenerate codon usage key	
Base	Bases
T	T
A	A
C	C
G	G
Y	C T
R	A G
S	G C
W	A T
K	T G
M	A C
B	C G T
D	A G T
H	A C T
V	A C G
N	A C G T

Table 1.1: Single letter code (base) used when referring to degenerate amino acid codons and their nucleotide bases.

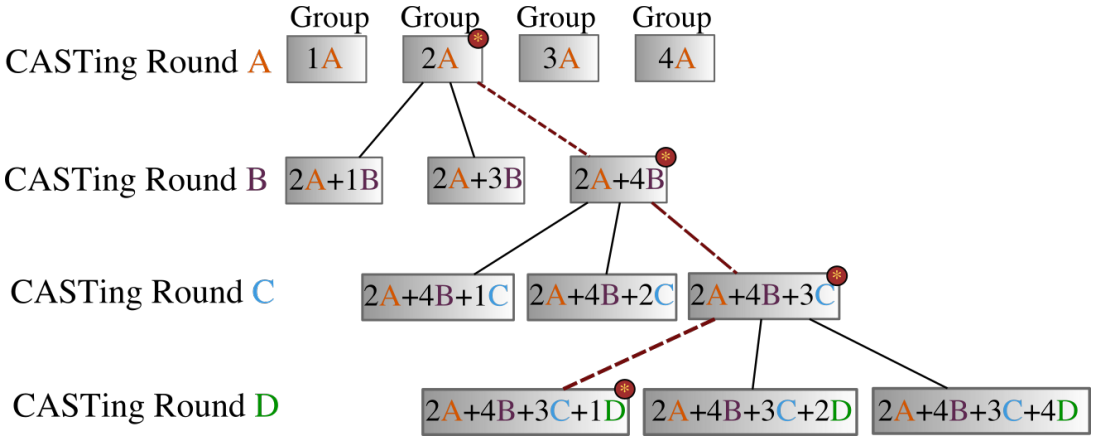


Figure 1.7: A hypothetical CASTing and ISM strategy. The route through the CASTing method is shown as a red dashed line with the group of the best variant for each round highlighted with a red circle. Four groups of amino acid residues are chosen (1-4) for targeted saturation with an appropriate degenerate codon. Each group must contain one or more amino acids. Round A of CASTing is performed on each group with an effective screen to identify desirable variants. The best variant (from group 2 in this case) is then used as the parent template for ISM with groups 1, 3 and 4. A further improved variant is identified with added group 4 substitutions. This iterative process is continued for rounds C and D, where each time an improved variant is identified.

By sampling the targeted residues in groups, rather than simultaneously, a significant reduction in screening effort is achieved. This is useful as we might expect to obtain an improved variant with the desired level of function over a much reduced time period. To highlight effectiveness of the CASTing strategy it is necessary to discuss the screening effort. In the example highlighted in Figure 1.7 we might imagine each group to contain two residues targeted for saturation with the NNK degenerate codon. This will incorporate all

amino acid variants with the lowest level of redundancy (32 codons for 20 amino acids). Screening effort can be calculated based on the number of targeted sites and the desired coverage of each library[3] using equation 1.1 [Patrick et al., 2003]. Where L is the required screening effort, F is the library completeness (99.9% = 0.99 (must be <1)) and V is the total number of possible codon combinations (32 x 32 for NNK), 4713 transformants would be required to screen 99% of each group assuming all codons have an equal probability of occurring.

$$L = -V \ln(1 - F) \tag{1.1}$$

The total screening effort requires sampling of 61,300 transformants. While still significant, this number is minuscule when considering it effectively samples a sequence space of $2.6\text{x}10^{10}$ combinants. If we were to reduce the library completeness to 95% and saturate all positions in the groups with the NDT degenerate codon (12 codons for 12 amino acids, see Section 4.1.1) this would reduce the total screening effort to 5590. This is a 75-fold reduction in screening effort at the expense of eight amino acids and 4% library coverage. From these examples it is clear to see how CASTing can be effectively applied to tackle the "numbers problem" [Reetz et al., 2008].

When contrasted to other directed evolution methods, CASTing has shown considerable improvements [Reetz et al., 2004, 2006]. In initial investigations, an expoxide hydrolase was targeted by directed evolution for improved stereoselectivity [Reetz et al., 2004] (Figure 1.8). The wild-type enzyme showed a 4.6-fold *S*-selectivity preference in the resolution of racemic glycidyl phenyl ether. Using epPCR and subsequent screening of 20,000 transformants, a variant was identified with an 11-fold *S*-selectivity preference [Reetz et al., 2004]. Using the CASTing approach, six groups of residues around the active site were targeted. With the same screening effort of 20,000 transformants, a 115-fold *S*-selective variant was obtained [Reetz et al., 2006].

[3] A CASTing tool (CASTer) is available for download from the Max-Planck Gesellschaft website (https://www.kofo.mpg.de/en/research/biocatalysis) and can be used to calculate screening effort.

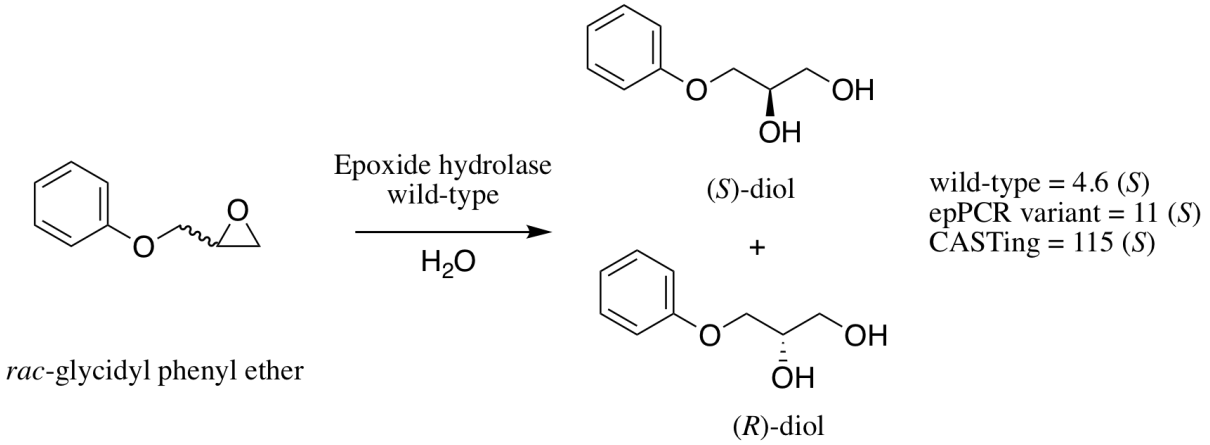


Figure 1.8: The hydrolytic cleavage of *rac*-glycidyle phenyl ether by an expoxide hydrolase. Reetz et al. [Reetz et al., 2006] used epPCR and CASTing to methods to separately engineer the same epoxide hydrolase. The CASTing variant demonstrated a 115-fold *S*-preference for the resolution of *rac*-glycidyle phenyl ether. By epPCR an 11-fold *S*-selective enzyme was identified; 10-fold lower than that engineered by CASTing .

1.2.3 Directed evolution assisted by computational design

With growing computational power, we are now in an era where *de novo* enzymes can be created entirely by *in silico* design [Kries et al., 2013]. The ability to design enzymes that catalyse non-natural reactions opens the door for an endless number of biological possibilities.

The first *de novo* protein to be predicted to a high degree of accuracy was for a novel protein fold called Top7 [Kuhlman et al., 2003]. Following production of the *in silico* model, the protein was expressed *in vivo*, purified and its structure solved. Astonishingly, this lay within 1.2 Å RMSD of the model [Kuhlman et al., 2003]. Top7 set the benchmark for future *de novo* design with functional enzymes soon appearing. A retro-aldolase [Jiang et al., 2008], kemp eliminase [Röthlisberger et al., 2008] and Diels-Alderase [Siegel et al., 2010] were all identified *in vitro* following generation of their *in silico* models. The computational algorithms used to create these enzymes are however not yet robust enough to credit levels of enzyme activity comparable with those in nature. Coupling *in silico* design with laboratory evolution has however demonstrated to be an effective method in achieving comparable levels [Khersonsky et al., 2010, Althoff et al., 2012]. In one example the *de novo* kemp eliminase [Röthlisberger et al., 2008] was further subject to directed evolution studies targeting improved catalytic efficiency (k_{cat}/K_m) [Blomberg et al., 2013]. The starting enzyme catalysed the cleavage of 5-nitrobenzisoxazole with a k_{cat}/K_m of 1,300 M^{-1}

s^{-1}. The enzyme was evolved through 17 rounds of directed evolution involving epPCR, DNA shuffling, and iterative cycles of CASTing (ISM). This produced a final 17 mutant variant with a k_{cat}/K_m of 230,000 M^{-1} s^{-1}. While this is not considered close to "enzyme perfection" [Knowles, 1991] it is comparable to natural enzyme catalytic efficiency. It should be noted that "enzyme perfection" is achieved when the rate determining step of the enzyme catalysed reaction is quicker than the diffusion limit (10^8-10^9 M^{-1} s^{-1}) as is observed with triosephosphate isomerase (TIM).

In another example a *de novo* aldolase [Jiang et al., 2008] was engineered to demonstrate natural enzyme like efficiency [Althoff et al., 2012, Giger et al., 2013, Obexer et al., 2017]. The enzyme was created *in silico* through careful placement of an active site catalytic lysine in a designed hydrophobic pocket. A well characterised TIM barrel scaffold was used and the final active enzyme was called retro-aldolase 95 (RA95). This catalysed the cleavage of *rac*-4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone (methadol) into acetone and 2-napthaldehyde (Figure 1.9) but with low efficiency ($k_{cat}/K_m = 0.19$ M^{-1} s^{-1}). The enzyme was evolved, first through 13 rounds of directed evolution consisting of epPCR and DNA shuffling with further focused saturation mutagenesis. Around 800 transformants from each round of evolution were screened using a microtitre plate assay. The screen used a fluorescence based detection method enabling activity of the most active variants to be identified. The final enzyme (RA95.5-8) displayed efficiency over 4000-fold that of the starting RA95 ($k_{cat}/K_m = 480$ M^{-1} s^{-1}). This enzyme was now also notably enantioselective with a 14-fold preference for *R*-methadol (RA95 = 2-fold selective for *S*-methadol) [Giger et al., 2013].

rac-4-hydroxy-4-(6-methoxy-2-naphthyl) -2-butanone

(methodol)

acetone

2-napthaldehyde

Figure 1.9: The retro-aldol cleavage of *rac*-4-hydroxy-4-(6-methoxy-2-naphthyl)-2-butanone (methadol) into acetone and 2-napthaldehyde by a *de novo* retro-aldolase (RA95).

Recently the emergence of ultra-high-throughput screening has enabled the constrained screening limit of classical microtitre assays to be dramatically increased [Yang and Withers, 2009, Guo et al., 2012]. Droplet-based microfluidic systems compartmentalise transformants and their reactions into individual aqueous droplets [Guo et al., 2012]. These droplets can be loaded with fluorogenic substrates and screened by fluorescence-activated droplet sorters (FADS) [Agresti et al., 2010, Yang and Withers, 2009]. It is reported that around 2000 droplets can be screened per second [Guo et al., 2012]. Obexer et al. [Obexer et al., 2017] have used this technology to further enhance the efficiency of RA95.5-8 for *R*-methadol. An efficiency of 34,000 M^{-1} s^{-1} (k_{cat}/K_m) was achieved for the best variant from screening 10^8 transformants. This enzyme, RA95.5-8F, also showed to be 480-fold selective for cleavage of *R*-methadol. The enzyme was also used to investigate the synthetically interesting C-C bond forming aldol reaction. This condensation reaction (Figure 1.9) had shown to be thermodynamically unfavourable for the synthesis of methadol ($K = 1.2\ M^{-1}$), however driving the reaction with excess acetone gave a conversion of 67% with an e.e. of 98.4% for *R*-methadol.

While the use of computation to create important biocatalysts has been demonstrated. It is clear that further advancements are required if we are to achieve synthetically useful enzymes without the need of additional laboratory evolution. Given these remarkable successes it seems appropriate to rework Figure 1.4 to factor in the capabilities of computational design in modern day directed evolution (Figure 1.10).

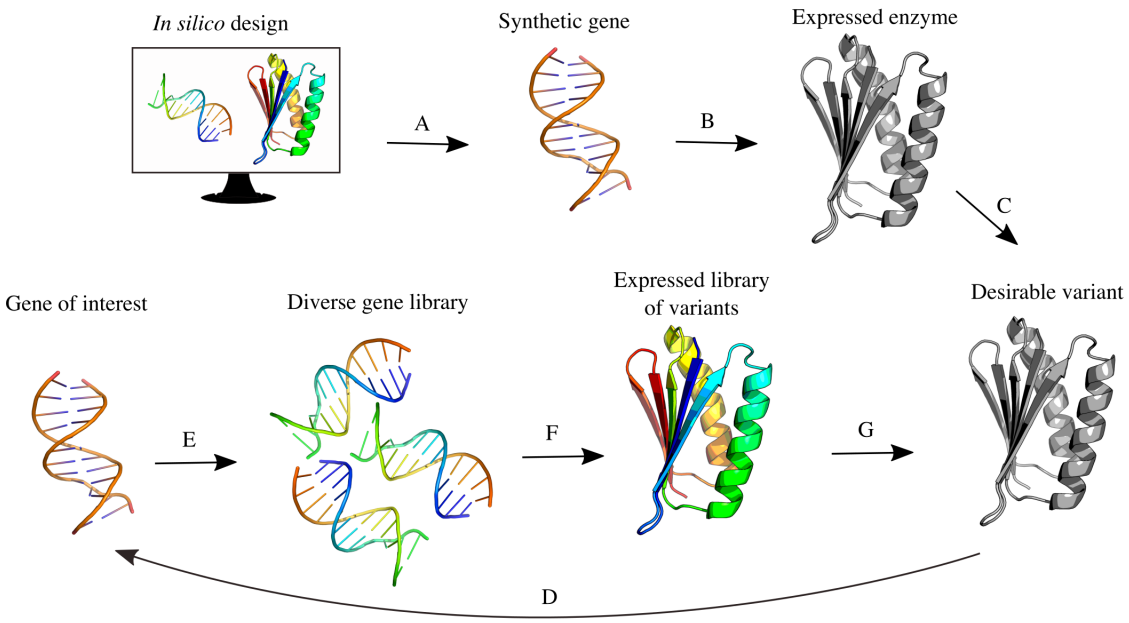


Figure 1.10: An appropriate schematic for modern day protein engineering where *in silico* enzyme design can be used to produce functional enzymes. When used in combination with a directed evolution strategy the best results are obtained [Blomberg et al., 2013, Savile et al., 2010, Giger et al., 2013]. A: The synthetic gene for the computationally predicted enzyme is synthesised and inserted into an appropriate host. B: Expression of the recombinant enzyme. C: The computationally designed enzyme can be characterised for a specified function. D: If the variant proves desirable, the diverse libraries of the gene can be generated (E) for subsequent expression (F) and screening. Iterations of the evolution process (steps D, E, F and G) can then be performed as many times as necessary.

1.3 Aldolases

In the previous sections we have given examples of aldolases as biocatalysts [Jennewein et al., 2006, Obexer et al., 2017]. They are of particular interest due to their ability to create new C-C bonds with excellent control over the formation of new stereogenic centers. The ability to create new C-C bonds allows for growing complexity to be introduced into organic molecules, a feature key to production of high value compounds [Gwon et al., 2012, Windle et al., 2014]. Aldolases are a family of specific lyases that catalyse the condensation reaction between an aldehyde acceptor and a ketone donor [Machajewski and Wong, 2000](Figure 1.11).

Unsurprisingly, one of the most prevalent ways nature achieves C-C bond formation is through the use of aldolases, given many natural metabolites exist as aldehyde and ketone moieties [Windle et al., 2014]. This family of enzymes can be divided into two classes

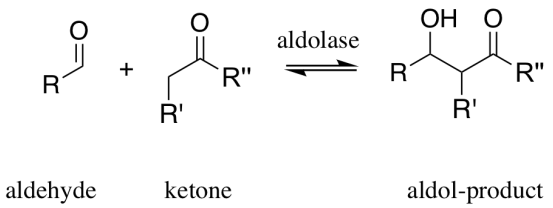


Figure 1.11: The reversible aldol condensation reaction between an aldehyde acceptor and a ketone donor.

based on their preference for activation of the donor molecule (Figure 1.12). The first of these, Class I aldolases, activate their donor molecule by formation of a Schiff base with a conserved active site lysine [Gefflaut et al., 1995]. Class II aldolases typically employ the use of a zinc ion cofactor for catalysis [Fessner et al., 1996]. A generalised mechanism for these enzymes is given in Figure 1.12.

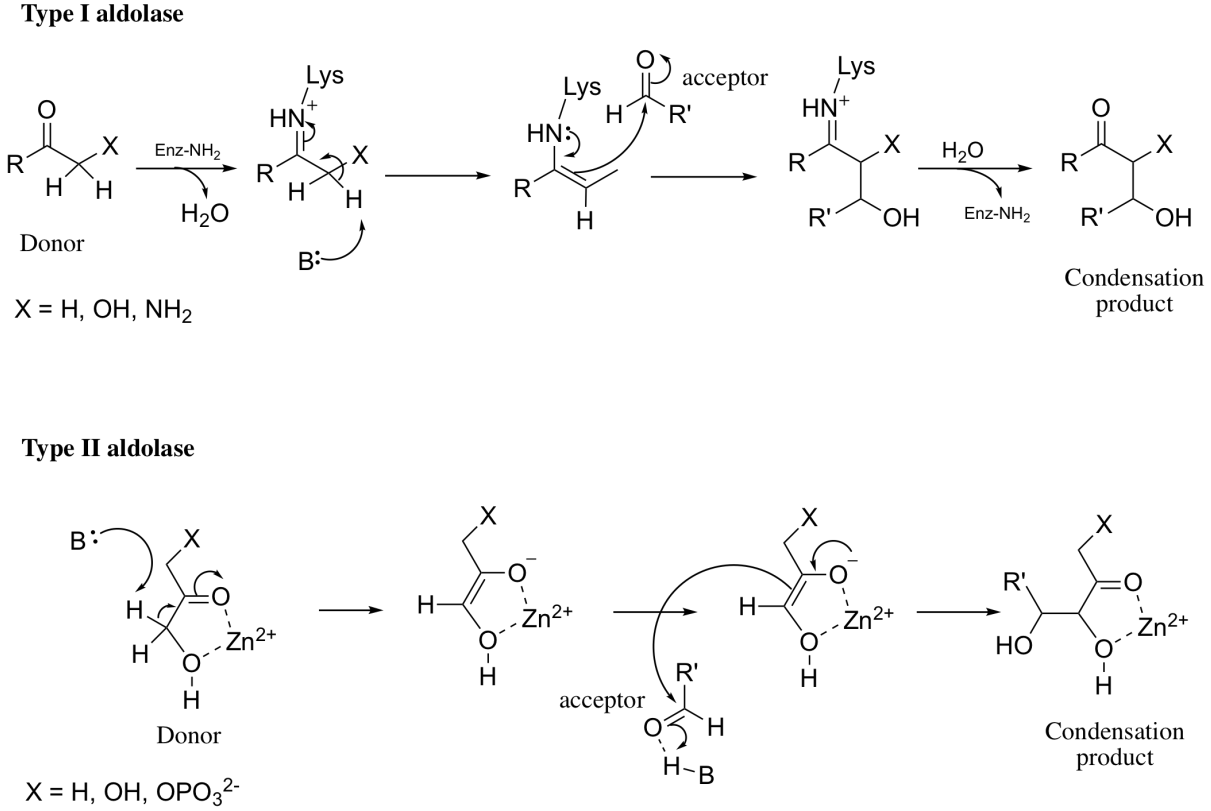


Figure 1.12: Generalised mechanisms for class I and class II aldolases. Type I aldolases: The donor molecule is activated by the formation of a Schiff base with a conserved active site lysine. Type II aldolases: The zinc-dependent reaction where deprotonation of the donor molecule results in the addition of the acceptor molecule to the enediolate forming the new C-C bond.

The use of non-native substrates in aldolases began to emerge after a publication by Wong and Whitesides [Wong and Whitesides, 1983] in the 1980s demonstrated the promiscuity

of fructose 1,6-bisphosphate aldolase in using a range of aldehyde acceptor molecules in catalysis. It has now long been shown that many enzymes of the aldolase family are capable of accepting a range of electrophilic aldehyde acceptor molecules to generate novel compounds [Windle et al., 2014]. Variability in the nucleophilic donor molecule however, has shown to be far less prevalent. For this reason Class I aldolases are further classified by their preference for the ketone donor. These include, but are not limited to, dihydroxyacetone or dihydroxyacetone phosphate, pyruvate or phospheneolpyruvate and glycine [Clapes et al., 2010]. Those active with glycine are more commonly referred as threonine aldolases (TA). These enzymes are of particular interest due to their ability to create two new adjacent stereocenters. These enzymes shall be discussed in the following section.

1.4 Threonine aldolases

The creation of two new stereogenic centers yields four new possible stereoisomers. This can be confusing given multiple nomenclatures are used when referring to different substrates with identical stereochemistries. To aid understanding it seems appropriate to summarise those used for particularly relevant TA compounds (threonine and phenylserine (PS))(Figure 1.13).

TAs catalyse the reversible cleavage of threonine into acetaldehyde and glycine. They are divided into two types based on their excellent enantiomeric preference for these substrates and are referred as L- or D-TAs (Figure 1.14). Within these types are further divisions depending on the stereoselective preference at the β-position (diastereoselectivity). L-*allo*-TA preferentially cleaves L-*allo*-threonine, L-TA preferentially cleaves L-threonine and *low-specificity* L-TA cleaves both L-threonine diasteromers. For D-TAs, only *low-specificity* enzymes are known. While this classification gives rise to multiple subdivisions of TA, the exact evolutionary function of these enzymes is not known. The active site pockets in all solved crystal structures, appears far to large to just accommodate threonine [Kielkopf and Burley, 2002, di Salvo et al., 2014, Uhl et al., 2015]. Further TA gene knockouts don't prove lethal in organisms [Liu et al., 1998]. Liu et al have reported that *Escherichia coli* (*E. coli)* TA (eTA) directed cleavage of L-threonine is not a prominent source of glycine within the

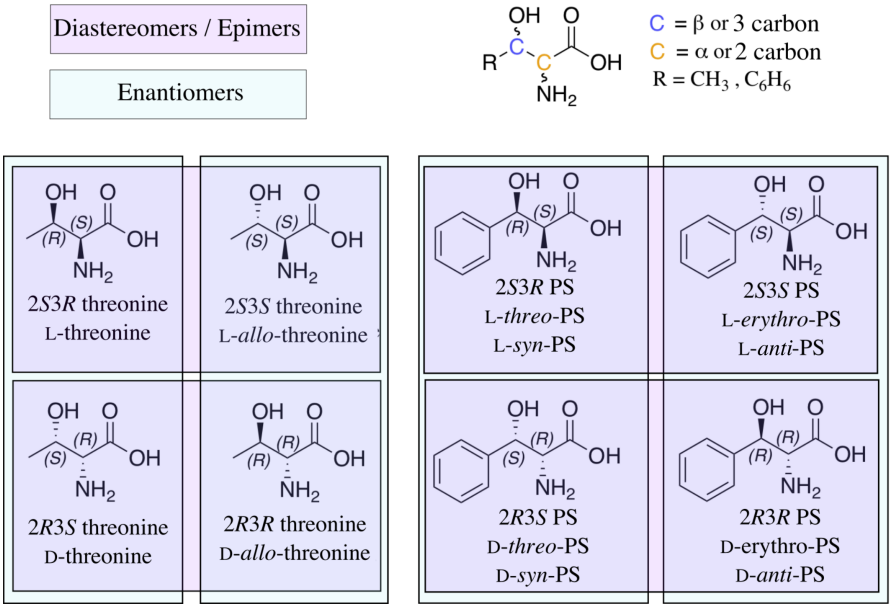


Figure 1.13: Structures of threonine, phenylserine (PS) and the stereochemical nomenclature used across the literature for each of their stereoisomers. Vertical green boxes are used to show pairs of enantiomers where the stereochemistry at both the α and β carbon is different. Horizontal pink boxes represent pairs of diastereomers/epimers of a single enantiomer where the stereochemistry at only the β position is different.

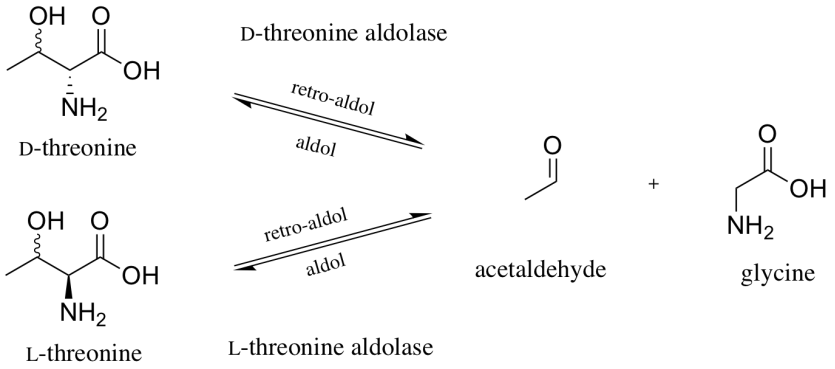


Figure 1.14: The reversible retro-aldol reaction of L- and D-threonine catalysed by L- and D- types of TA.

cell. If the main pathway via serine hydroxymethyltransferase is blocked it may however suffice as an alternative route [Liu et al., 1998]. It was shown that 4-phosphoerythronate dehydrogenase deficient *E. coli* strains required L-TA for synthesis of pyridoxal 5'-phosphate (PLP). The enzyme is capable of catalysing the condensation of glycoaldehyde and glycine to form L-4-hydroxythreonine. Once phosphorylated to L-4-phosphohydroxythreonine, this compound is used as an intermediate in the PLP synthesis pathway [Kim et al., 2010]. Suitably, PLP is also an essential cofactor in TA catalysed reactions [Liu et al., 1997].

L- and D-TAs are evolutionarily and structurally distinct. L-TAs belong to the aspartate amino

transferase fold family and D-TAs to the alanine racemase family [Paiardini et al., 2003]. They classify as group 1 members with respect to the classification of enantiocomplementary enzymes. This includes enzymes with different folds and mirror-image active sites [Mugford et al., 2008, Uhl et al., 2015]. Unlike L-TAs, D-TAs require divalent metal ions for function [Liu et al., 2000]. The first crystal structure of a D-TA from *Alcaligenes xylosoxidans* (*A. xylosoxidans*) showed a well coordinated manganese ion within the active site (Figure 1.15A and B).

Several high-resolution structures have been solved for L-TAs, these include the *low-specificity* L-enzymes from *E. coli* (Section 5.2.2) [di Salvo et al., 2014] and *Thermotoga maritima* [Kielkopf and Burley, 2002] as well as an L-*allo*-enzyme from *Aeromonas jandaei* (*A. jandaei)* [Qin et al., 2014]. The structures of these enzymes all exhibit identical protein folds which overlay perfectly despite only having 36% sequence identity [4] (Figure 1.15C). Furthermore the active sites of the enzymes are virtually indistinguishable (Figure 1.15D). Docking experiments were performed on the D-TA in order to determine a likely substrate binding pose. The enzyme was docked with 2*R*3*R* and 2*R*3*S* PS with suitable poses suggesting a histidine to be involved in proton abstraction through a coordinated water molecule. Further it appeared that the manganese ion likely functions as a key Lewis acid in the mechanism (Figure 1.16). To date no mutagenesis studies have been performed on D-TAs. These will be necessary to implicate the role of individual residues in catalysis.

[4]Protein sequences were aligned for *low-specificity* L-TAs from *E. coli* (gene ItaE) and *T. maritima* (gene Tmari1752) and L-*allo*-TA from *A. jandaei* (gene ItaA) using the Clustal Omega program (available at http://www.clustal.org/omega)

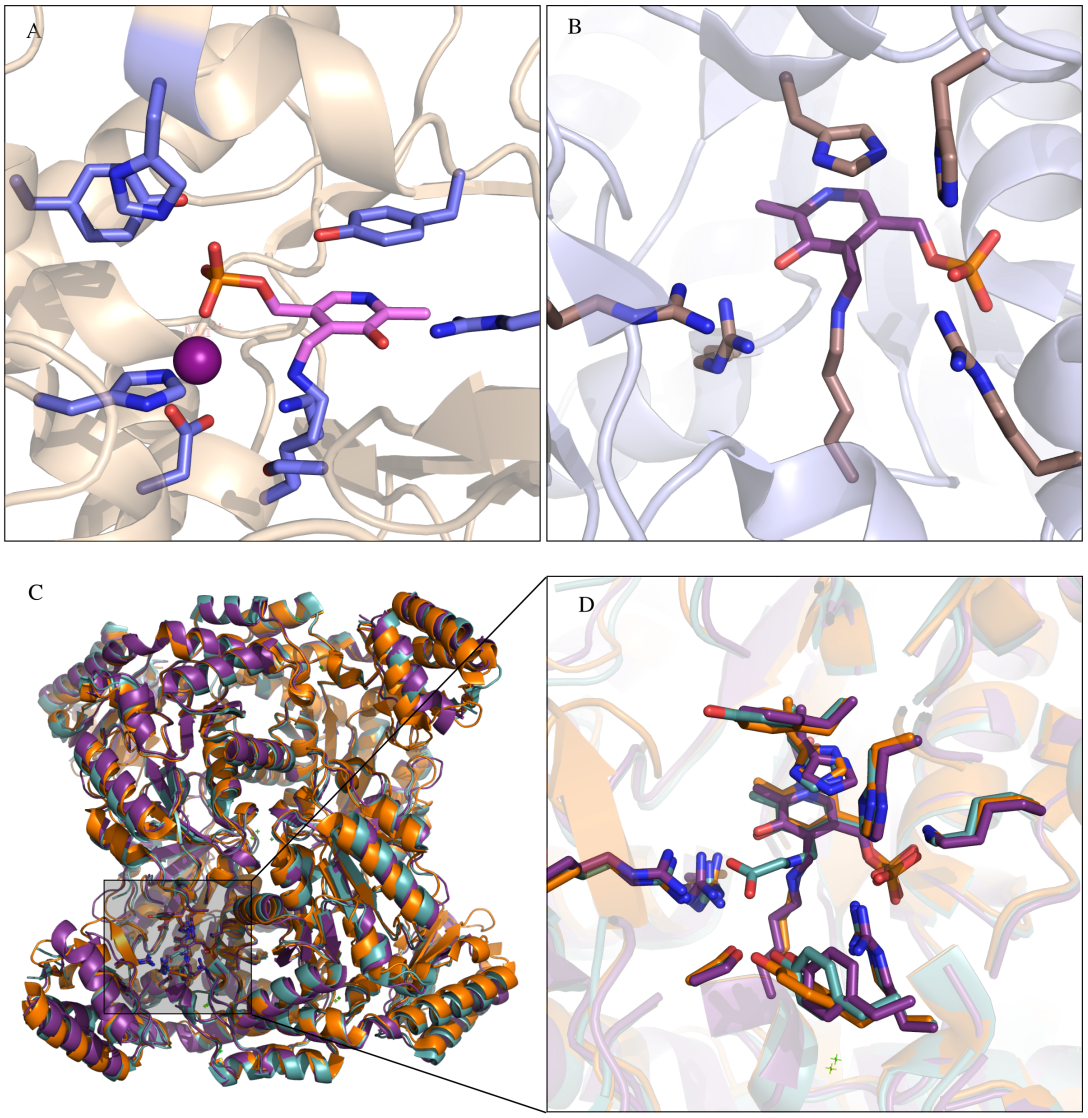


Figure 1.15: A: The active site of the first crystal structure of a D-TA from *A. xylosoxidans* (PDB file 4V15) [Uhl et al., 2015]. A coordinated manganese ion (purple sphere) is in close proximity to the PLP cofactor (pink) forming a Schiff base with a lysine residue. A stacking interaction between an active site tyrosine is observed with the pyridine ring of PLP. B: The active site of an L-TA from *T. maritima* (PDB file 1M6S) [Kielkopf and Burley, 2002]. The PLP cofactor (purple) forms a Schiff base to an active site lysine with a stacking interaction forming between a histidine and the pyridine ring of PLP. C: Three tetrameric L-TA structures, *low-specificity* L-TAs from *E. coli* (purple) and *T. maritima* (orange) and L-*allo*-TA from *A. jandaei* (blue) aligned by their mainchain Cα-atoms. Despite having low protein sequence identity (36%) high structural similarity is observed for the enzymes. D: Expanded view of the active site for enzymes shown in C. Residues and their orientations show to be virtually identical. Both *E. coli* and *T.maritima* TAs show covalently bound PLP to lysine residues. The *A. jandaei* enzyme shows covalently bound ligand (glycine) to the PLP cofactor.

In contrast L-TAs have been subject to a considerable number of mutagenesis studies [Gwon and Baik, 2010, Giger et al., 2012, di Salvo et al., 2014, Fesko et al., 2018]. It was initially

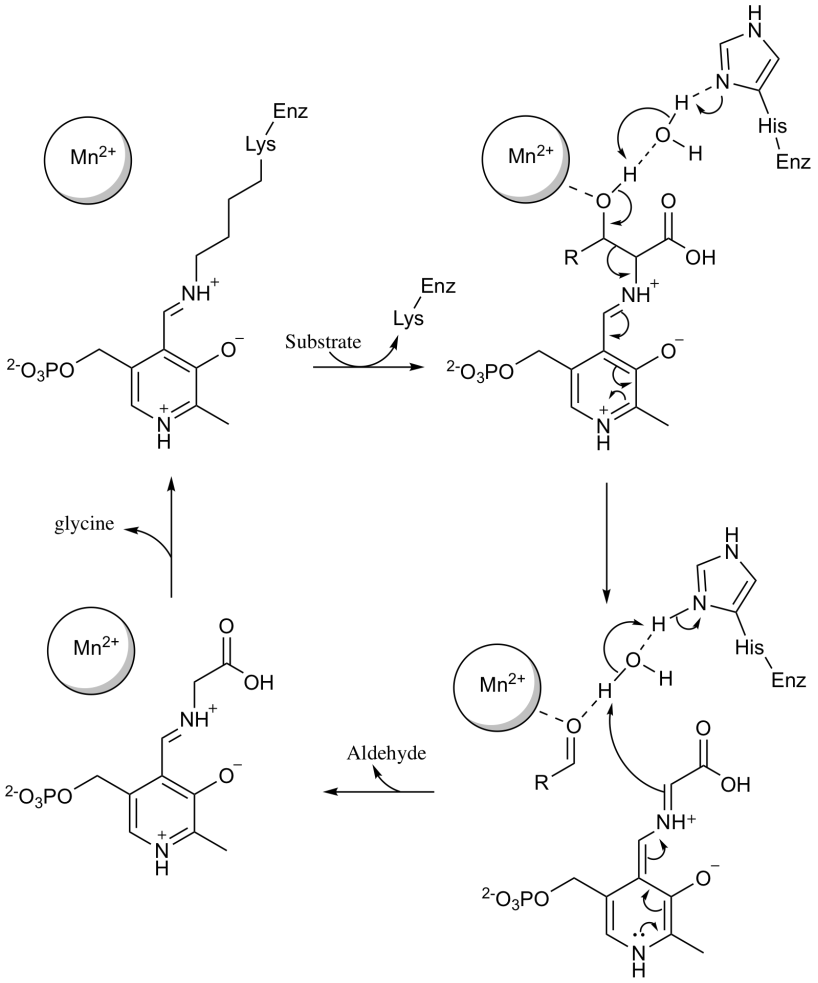


Figure 1.16: A proposed mechanism for D-TA retro-aldol cleavage. Upon addition of substrate a transaldimation reaction occurs and PLP binds to the substrate through a Schiff base. The positively charged manganese ion coordinates the bound substrate through interaction with the negatively charged aldol oxygen. A histidine residue functions as the general base abstracting the hydroxyl proton through a water molecule. The C-C bond is cleaved yielding a resonance stabilised aldimine (PLP-glycine) and free aldehyde. Reprotonation then occurs through the histidine-water relay. A further transaldimation allows release of glycine and reforms the PLP-enzyme linkage.

proposed that an active site histidine directly mediated proton abstraction in the retro-aldol cleavage mechanism [Giger et al., 2012]. More recently, di Salvo et al. [di Salvo et al., 2014] have shown this to be unlikely. Mutagenesis of the proposed histidine to a non-basic residue still yielded activity and further, threonine bound crystal structures revealed a coordinated water molecule in place for proton abstraction. A mechanism for retro-aldol cleavage by L-TA is shown in Figure 1.17.

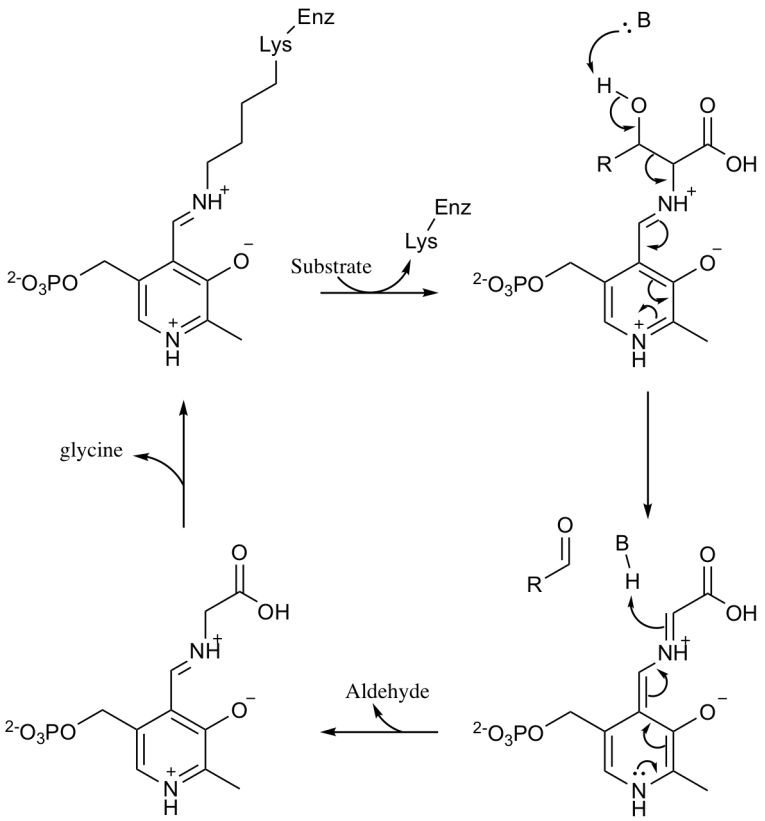


Figure 1.17: A generalised mechanism for L-TA retro-aldol cleavage. Through addition of substrate a transaldimation reaction occurs and PLP binds to the substrate through a Schiff base linkage. C-C bond cleavage occurs as the hydroxyl proton of the substrate is abstracted by a nucleophilic base. This yields the resonance stabilised aldimine (PLP-glycine) and free aldehyde. Reprotonation ensues and glycine is released through transaldimation with the appropriate active site lysine.

Synthetically, TAs are of interest due to their high enantioselectivity and promiscuity in accepting an extensive range of aldehydes [Steinreiber et al., 2007, Chen et al., 2017, Beaudoin et al., 2018]. Through condensation with glycine, TAs can synthesise single enantiomer β-hydroxy-α-amino acid products. These compounds are extremely important to pharmaceutical and agrochemical industries [Patel, 2013]. Antibiotics including chloramphenicol, thiamphenicol and vancomycin all exibit these key structural units in their active structures. Further, β-hydroxy-α-amino acids can be used directly as the active pharmaceutical ingredient (API), for example L-DOPS an active Parkinson's disease drug. Subsequently, routes to these compounds are required. TAs offer such a route however they are restricted in their applications due to poor selectivity at the β-position resulting in a mix of diasteromers and further, poor yields. Notably, TAs have also been used for their excellent enantioselectivity in the resolution of racemic β-hydroxy-α-amino acids [Herbert et al., 1994].

1.5 Main Aims

The main aims of this thesis are to engineer eTA to exhibit improved stereoselectivity towards a single diasteromer of PS. These efforts will be divided across three results chapters with their aims detailed below.

1.5.1 To clone the gene, express and purify the enzyme and develop a high-throughput screening assay

In the first results chapter, this thesis will focus on cloning the eTA gene and subsequent expression and purification of the enzyme. Particular attention will also be given to development of a high-throughput screen. This screen will be used to identify variants with improved PS stereoselectivity through spectrophotometric detection of benzaldehyde.

1.5.2 To engineer eTA by directed evolution

In the second results chapter efforts will be directed to the screening of focussed libraries. We aim to screen pairs of residues around the active site of eTA by the method of CASTing and ISM [Reetz et al., 2005, 2006]. This will aim to identify synergistic interactions that would otherwise be unidentifiable by single site saturation mutagenesis. We will use the results of CASTing and ISM to make rational combinations of variants and characterise their stereoselectivity using Michaelis-Menten kinetics.

1.5.3 To obtain structural insight into the poor stereoselective mechanism of eTA

In the third and final results chapter, this thesis will use X-ray crystallography to obtain a high resolution structure of eTA. With this structure we aim to gain insight into the poor stereoselectivity observed in the enzyme for PS. For this we will use molecular mechanics (MM) to investigate binding of pre-3*R* and pre-3*S* orientations of benzaldehyde.

Further, more robust quantum mechanics/molecular mechanics (QM/MM) simulations will investigate the mechanism of the stereogenic step in reaction.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Technical equipment

Equipment	Manufacturer
Avanti J-26 XP centrifuge	Beckman Coulter, Brea, California, USA
Micro Centaur	MSE, London, UK
Microplate well dispenser	Genetix, UK
Gallenkamp Economy Incubator SIZE 1	Sanyo, Watford, UK
Microplate Shaker	JENCONS, UK
Orbisafe Orbital Incubator	Sanyo, Watford, UK
MS1 Minishaker	IKA, Staufen, Germany
Vortex Genie 2	Scientific Industries, New York, USA
Vari-Gel midi system	CamLab, Cambridge, UK
Mini-PROTEAN Electrophoresis system	Bio-Rad Laboratories, Hertfordshire, UK
PowerPac Basic power supply	Bio-Rad Laboratories, Hertfordshire, UK
AKTAprime plus	GE healthcare, Little Chalfont, UK
Supadex S200 26/60 gel filtration column	Little Chalfont, UK
Uvicon 930 spectrophotometer	Kontron Instruments, UK
Nanodrop 2000c spectrophotometer	Thermo Fisher Scientific, Massachusetts, US
FLUOstar Galaxy plate reader	BMG Labtech, Ayelsbury, Bucks, UK
Jenway 3020 pH meter	Bibby Scientific, Stone, UK
PTC-100 Programmable Thermal Controller	GMI-Inc, Minneapolis, US
Grant JB1 Waterbath	Grant Instruments, Shepreth, UK

Table 2.1: Equipment, and manufacturer, used in this thesis.

Chemical	**Manufacturer**
Agarose	Melford Laboratories, Suffolk, UK
Ammonium Acetate	Fisher Scientific, Loughborough, UK
Agar	Melford Laboratories, Suffolk, UK
Ammonium persulphate	Acros Organics, (Fisher Scientific)
Acrylamide, 30%	Severn Biotech, Kiddiminster, UK
Ampicillin	Formedium, Norfolk, UK
Benzaldehyde	Acros Organics, (Fisher Scientific)
Bromophenol Blue	Sigma-Aldrich, Dorset, UK
Dimethyl sulphoxide (DMSO)	Sigma-Aldrich, Dorset, UK
DNase	Sigma-Aldrich, Dorset, UK
Ethanol	Fisher Scientific, Loughborough, UK
Ethylenediamine tetra acetic acid (EDTA)	Sigma-Aldrich, Dorset, UK
Glycerol	Fisher Scientific, Loughborough, UK
Glycine	Fisher Scientific, Loughborough, UK
Instablue	BDH Biochemical, Poole, UK
Isopropylβ-D-1-thiogalactopyranoside (IPTG)	Generon, Berkshire, UK
Imidazole	Acros Organics (Fisher Scientific)
Lysozyme	Sigma-Aldrich, Dorset, UK
β-mercaptoethanol	Stratagene, Cambridge, UK
L-*threo*-phenylserine (2*S*3*R*)	Santa Cruz Biotech, US
L-*erythro*-phenylserine (2*S*3*S*)	NetChem, New Brunswick, CA
Polyethylene glycol 400	Sigma-Aldrich, Dorset, UK
Propan-2-ol	Fisher Scientific, Loughborough, UK
Pyrodoxal 5'-phosphate	Sigma-Aldrich, Dorset, UK
Sodium chloride	Fisher Scientific, Loughborough, UK
Sodium dodecyl suplhate (SDS)	Sigma-Aldrich, Dorset, UK
Sodium hydroxide (NaOH)	Fisher Scientific, Loughborough, UK
Sodium phosphate (dibasic)	Sigma-Aldrich, Dorset, UK
Sodium phosphate (monobasic)	Sigma-Aldrich, Dorset, UK
Tetramethlethylediamine (TEMED)	Sigma-Aldrich, Dorset, UK
Tris base	Sigma-Aldrich, Dorset, UK
Tryptone	Sigma-Aldrich, Dorset, UK
Yeast extract	Sigma-Aldrich, Dorset, UK

Table 2.2: Chemicals, and their manufacturer, used in this thesis.

2.1.2 Media

Unless otherwise stated all media used in this thesis was 2 TY. 1 L of this media contained 16 g tryptone, 10 g yeast extract, 5g NaCl and autoclaved at 121^{o}C for 20 minutes. If solid phase 1.5% (w/v) agar was supplemented prior to autoclaving.

2.2 General methods

2.2.1 Centrifugation

Samples less than 2 mL were centrifuged at the appropriate rpm using a Micro Centaur bench top centrifuge at room temperature unless stated otherwise. Those samples greater than 2 mL were centrifuged using an Avanti J-26 XP at 4 ^{o}C unless stated otherwise.

2.2.2 pH measurements

pH measurements were performed at room temperature using a Jenway 3020 pH meter, calibrated according to the manufacturer's guidelines.

2.2.3 Antibiotic stocks

Ampicillin and tetracyclin were used at a working concentration of 50 μg mL^{-1} in both solid and liquid phase media where specified. Stocks were made at 50 mg mL^{-1} and stored at -20^{o}C.

2.2.4 Culture growth

Single colonies were picked, using a sterile tip, from 2 TY plates supplemented with 50 μg mL^{-1} of the necessary antibiotic and used to inoculate 5 mL of 2 TY media containing 50 μg mL^{-1} of the necessary antibiotic. Cultures were grown over night for 16 hours at 37^{o}C, with shaking at 200 rpm in an orbital incubator. For expressions of 1 L: 50 μL of starter culture was used to inoculate a 50 mL day culture and incubated for 8 hours at 37^{o}C, with shaking at 200 rpm in a 250 mL conical flask. 10 mL of day culture was used to inoculate 1 L of 2 TY supplemented with 50 μg mL^{-1} ampicillin in a 2.5 L conical flask. Cultures were incubated at 37^{o}C, with shaking at 200 rpm until an optical density (O.D) of 0.6 at 600 nm was reached. Protein expression was induced by addition of IPTG to a final concentration of 0.1 mM. For 50 mL expressions: 50 μL of the starter culture was used to innoculate 50

mL of 2 TY containing 50 μg mL^{-1} ampicillin in a 250 mL conical flask. Cultures were incubated at 37^{o}C, with shaking at 200 rpm until an optical density (O.D) of 0.6 at 600 nm was reached. Protein expression was induced by addition of IPTG to a final concentration of 0.1 mM.

2.2.5 Glycerol stocks

Glycerol stocks were made from starter cultures (Section 2.2.4). A 50% (v/v) stock of bacterial culture was created by adding 250 μL of culture to 250 μL of sterile glycerol in a sterile Nunc CryoTube and stored at -80^{o}C.

2.3 DNA methods

2.3.1 Bacterial strains

Bacterial strains used in DNA manipulations were:
E. coli XL1-Blue (Agilent Technologies, Cheshire, UK)
*rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *sup*E44 *rel*A1 *lac* [F' *pro*AB *lacI*^{q}ZΔM15 *Tn10* (Tet^r)]
E. coli XL10-Gold (Agilent Technologies, Cheshire, UK)
Tet^r Δ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrr*)173 *end*A1 *sup*E44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 *lac* Hte
[F' proAB *lacI*^{q} ZΔM15 *Tn*10(Tet^r)Amy Cam^r]
E. coli BL21 (Agilent Technologies, Cheshire, UK)
(B F$^-$ *ompT* *hsdS*($r_B^-m_B^-$) *dcm*$^+$ Tet^r *endA* Hte

2.3.2 Cloning the gene for *E. coli* threonine aldolase into the pKK223-3 expression vector

The codon optimised gene for *low-specificity* L-threonine aldolase from *E. coli* (ItaE), was previously ordered from Genescript (New Jersey, USA) pre-cloned into a pUC57 vector (pUC57-ItaE). The pKK223-3 plasmid was provided by Professor Alan Berry

and used to create the pKK-ItaE construct for expression of *low-specificity* L-threonine aldolase from *E. coli* (eTA) (see Figure 3.3B). The *fast cloning* method described by Li et al. [Li et al., 2011] was used with sets of primers designed to amplify both ItaE and pKK223-3 in a way that allowed for recombination. Two reactions were set up to amplify both the gene and plasmid separately in 0.5 mL thin walled PCR tubes. Reactions contained 5 x Q5 buffer (10 μL), 2 μM dNTPs (5 μL), nuclease free water (29 μL), Q5 DNA polymerase (0.5 μL) with either pUC57-ItaE (0.5 μL) or pKK223-3 (0.5 μL). Reactions also included, for amplification of ItaE: 10 μM forward primer (5'-ACAGGAAACAGAATTCATGGAACATCACCATCACCACCAC-3') (2.5 μL) and 10 μM reverse primer (5'-AGAAGCTTGGCTGCAGTTATTAACGCGCCAGAAACGCACG-3') (2.5 μL) or for amplification of pKK223-3: 10 μM forward primer (5'-CTGCAGCCAAGCTTCTGTTTTG-3') (2.5 μL) and 10 μM reverse primer (5'-GAATTCTGTTTCCTGTGTAAATTGTTATCC-3') (2.5 μL).

The conditions for the PCR reaction were as follows. For ItaE: 98^{o}C (30 s) followed by 20 cycles of 98^{o}C (10 s), 71 ^{o}C (20 s), 72 ^{o}C (30 s). A final extension step at 72^{o}C (120 s) completed the reaction which was then kept on ice at 4 ^{o}C. For pKK223-3: 98^{o}C (30 s) followed by 20 cycles of 98^{o}C (10 s), 65^{o}C (20 s), 72 ^{o}C (150 s). A final extension step at 72^{o}C (120 s) completed the reaction which was then kept on ice at 4 ^{o}C.

45 μL of each reaction was combined in a sterile 0.5 mL thin walled PCR tube with 1 μL of DpnI enzyme and incubated for 1 hour at 37^{o}C. 2.5 μL of the reaction mixture was then transformed into *E. coli* XL10-Gold cells (Section 2.3.5).

2.3.3 Plasmid purification

5 mL cultures of 2 TY supplemented with 50 μg mL^{-1} ampicillin were inoculated with single colonies of 5 μL of a glycerol stock and grown for 16 hours with shaking at 200rpm in an orbital shaker. Plasmid DNA was then purified from these starter cultures using a Wizard Plus SV Minipreps DNA Purification System (Promega, Southampton, UK) according to the manufacturer's guidelines. DNA quantification was performed using a Nanodrop2000.

2.3.4 Agarose gel electrophoresis

Agarose gel electrophoresis was performed with 1% (w/v) agarose disolved in TAE buffer. A 1 kb DNA ladder was used for determining fragment sizes.

10 x TAE buffer
10 mM EDTA
200 mM glacial acetic acid
400 mM tris base

2.3.5 Transformation

Transformations were carried out using the heat shock method according to the method provided with the cells used. Plasmid DNA was transformed into either *E. coli* XL1-Blue *E. coli* XL10-Gold and *E. coli* BL21-Gold cells as specified in each section.

2.4 Protein methods

2.4.1 Protein purification

E. coli cells containing the His_6-tagged eTA or variant were cultured as in Section 2.2.4 and purified using Chelating Sepharose fast flow resin, chelated with nickel ions in a batch purification method.

For 1 L expressions: Cell pellet was harvested from 1 L of bacterial culture by centrifugation at 12 000 g for 20 min. The cell pellet was re-suspended in 50 mL of wash buffer per 10 g of cells using a manual homogeniser. The re-suspended cells were lysed at 20 kpsi using a Cell Disruptor from Constant Cell Disruption Systems (Northamptonshire, UK). Insoluble cell debris was separated from the soluble protein by centrifugation at 40000 g for 45 min at 4^{o}C. A fraction of the crude soluble lysate supernatant was analysed by SDS PAGE. This supernatant, containing protein fraction, was then loaded onto 5 mL of Chelating Sepharose Fast Flow resin in a 50 mL Falcon tube, and incubated for 30 min with agitation at 4^{o}C.

Centrifugation at 4000 g for 5 min at 4^{o}C produced a supernatant containing any non-bound proteins, a fraction of this supernatant was analysed by SDS PAGE as a wash sample. The resin was then thoroughly re-suspended using 40 mL of washing buffer and centrifuged at 4000 g for 5 min at 4^{o}C after which the supernatant was discarded, this step was repeated three times. Elution of the His6-tagged protein was achieved by addition of 20 mL elution buffer and incubation for 1 hour at 4^{o}C with agitation. After centrifugation at 4000 g for 5 min at 4^{o}C the supernatant containing the His_6- tagged protein was collected.

For 50 mL expressions (all CASTing variants): Cell pellets were harvested at 6000 g for 20 min and stored at -20^{o}C for 1 hour or until required. The cell pellet was then re-suspended in 2 mL of wash buffer. The re-suspended cells were lysed by sonication. Re-suspended cells were kept on ice and subject to 3 rounds of sonication at 40% amplitude for 30 seconds with 30 seconds resting on ice in between each round. Insoluble cell debris was separated from the soluble protein by centrifugation at 12000 g for 1 hour at 4^{o}C using a Micro Centaur benchtop centrifuge. A fraction of the crude soluble lysate supernatant was analysed by SDS PAGE. This supernatant, containing protein fraction, was then loaded onto an equilibrated gravity flow column containing 2 mL of Chelating Sepharose Fast Flow resin and incubated for 30 min with agitation. Fractions of the unbound flow through were collected and analysed by SDS PAGE. 40 mL (4 column volumes) of wash buffer were added to the column and after flow though after each 10 mL analysed by SDS PAGE. The resin was then thoroughly re-suspended using 2 mL of elution buffer and left for 30 minutes. The His_6-tagged protein was then collected.

Wash buffer

50 mM Tris/HCl pH 7.4

20 mM Imadazole

500 mM NaCl

Elution buffer 50 mM Tris/HCl pH 7.4

500 mM Imadazole

500 mM NaCl

2 x loading buffer for SDS PAGE:

2 mL 10% (w/v) SDS

200 μL 0.2% (w/v) bromophenol blue in ethanol

125 mg DTT

1 mL glycerol

170 μL 1 M Tris/HCl pH 6.9

163 μL H_2O

2.4.2 Protein dialysis

Dialysis of proteins was performed using BioDesign dialysis tubing (14,000 Da molecular weight cut off). Proteins were dialysed into the appropriate buffer at 500 times the volume of the protein. Dialysis was performed with stirring at 4^{o}C for 6 hours then a fresh buffer was used for a final 6 hours of dialysis. Or using a 2 mL Slide-A-Lyzer MINI Dialysis Device, 20,000 Da molecular weight cut off (Thermo Fisher Scientific (Life Technologies)).

2.4.3 Protein concentration determination

Protein concentration was determined using a Nanodrop 2000. Values for the molar co-extinction coefficient were input and the molecular weight of the protein. 2 μL of purified

protein was added to the pedestal and three repeats of the measurement were performed. The final concentration was determined as the average of the three repeats.

2.4.4 Increasing protein concentration

Protein samples were concentrated using a Vivaspin Protein Concentrator (10 kDa molecular weight cut off) according to the manufacturer's guidelines.

2.4.5 Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) was performed using pre-cast Mini-PROTEAN Gels, 8-16% (Bio-Rad Laboratories) or in home made gels in resolving gel (15% (v/v) acrylamide) and stacking gel (3.75 % (v/v) acrylamide). Gels were made as described below:

Resolving gel:
7500 μL 30% (w/v) acrylamide
3750 μL 1.5 M Tris/HCl pH 8.8
3500 μL dH_2O
150 μL 10% (w/v) SDS
50 μL 25% (w/v) APS
5 μL TEMED

Stacking gel
625 μL 30% (w/v) acrylamide
625 μL 1.0 M Tris/HCl pH 6.9
3650 μL dH_2O
50 μL 10% (w/v) SDS
50 μL 25% (w/v) APS
5 μL TEMED

Running buffer

14.4 g glycine

3 g Tris base

1 g SDS

140 g β-mercaptoethanol

dH_2O made up to 1 L

2.4.6 Size exclusion chromatography

Size exclusion chromatography was performed to obtain the purest fraction possible for crystallographic analysis (Section 5.2). An AKTA Prime purification system (GE Healthcare Life Sciences) was used with a 16/60 Superdex S200 column equilibrated with with degassed 50 mM sodium phosphate buffer, pH 8.0. Purified protein was injected onto the column at 10 mg mL^{-1}. 2 mL fractions were collected at a flow rate of 2 mL min^{-1}. The main peak of eTA eluted after approximately 50 mL (Figure 2.1). Collected fractions were analysed by SDS PAGE with the most concentrated of those pooled. eTA was then concentrated to 10 mg mL^{-1} by as per Section 2.4.4.

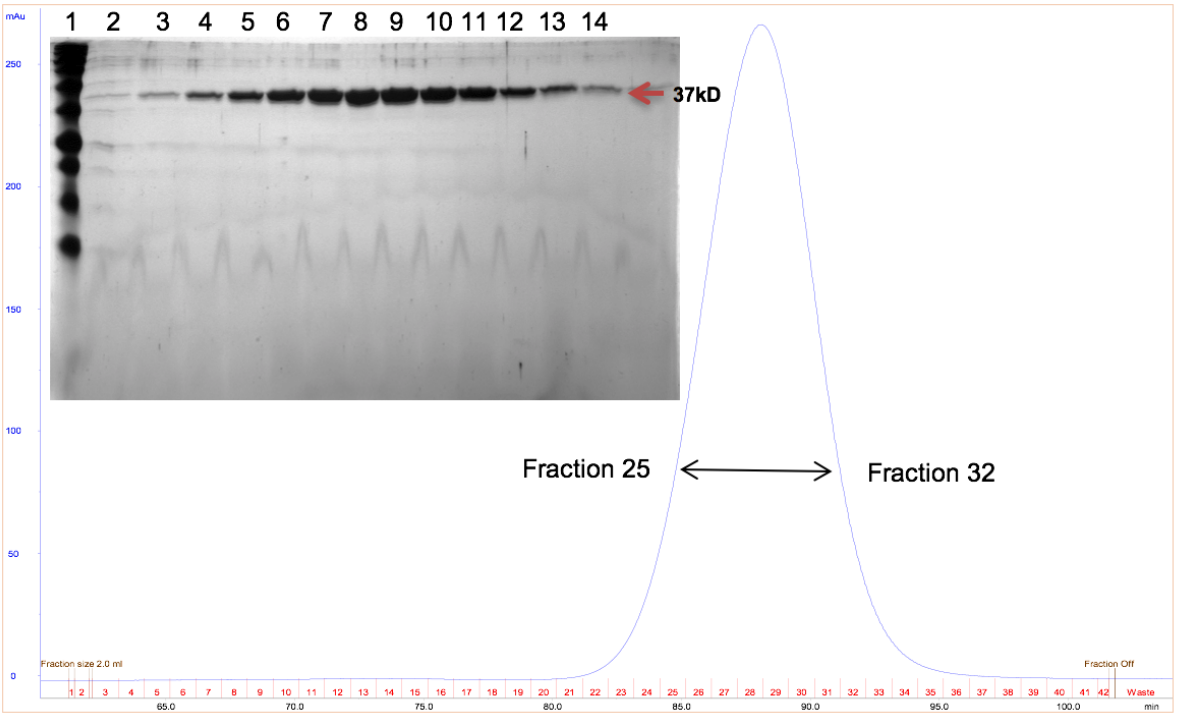


Figure 2.1: Gel filtration trace of concentrated eTA at 280nm. Also shown: SDS PAGE gel of collected fractions. Lane 1). Molecular weight marker. Lanes 2-15). Collected eTA fractions. Lanes 5-12). Fractions 25-32 from gel filtration trace.

2.5 Assay and screening methods

2.5.1 Wavelength absorbance measurements.

Wavelength reaction spectra were obtained on a Nanodrop 2000 using the described conditions. Reactions were set up on a 0.5 mL scale. For each measurements 2 μL of the reaction was placed onto the pedestal with instant measurement. The sample was then wiped off using a dust free wipe. 2 μL of the same reaction was then re-added and a measurement taken. The process was repeated as necessary.

2.5.2 Library generation and site-directed mutagenesis

CASTing libraries for groups 1 ($H83_{NDT}$ $F87_{NDT}$, 2 ($H126_{NDT}$ $F127_{NDT}$) and 3 ($A81_{NDT}$ $L86_{NDT}$) were generated by PCR using designed mutagenic primers (Appendix Table A.2) and a QuikChange lightning mutagenesis kit (Agilent Technologies, Cheshire, UK). Reactions contained: 10 x reaction buffer, nuclease free H_2O, 2 mM dNTPs, 100 ng of the

respective forward and reverse primers, 50 ng of parent template DNA and pfu turbo DNA polymerase. The PCR reactions were as follows: 95^{o}C (30 s) followed by 17 cycles of 95^{o}C (30 s), 55^{o}C (60 s), 68^{o}C (336 s). Reactions were then subject to Dpn I digest for 1 hour at 37^{o}C. 3 μL of this was then transformed into 100 μL of *E. coli* XL10-Gold as per Section 2.3.5. By this process roughly 1000 colonies were achieved for each library.

Site-directed mutagenesis for rational combinations of variants was performed as described above using the respective primers from Appendix Table A.2.

For group 4 ($Y30_{NDT}$ $R229_{NDT}$) libraries were generated by PCR using designed mutagenic primers (Appendix Table A.2and QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent Technologies, Cheshire, UK). Reactions contained 10 x reaction buffer, nuclease free H_2O, 2 mM dNTPs, 100 ng of each of the respective forward primers, 50 ng of parent template DNA, Quik solution (2% (v/v)) and QuikChange Lightning Multi enzyme blend. The PCR reaction was performed as follows: 95^{o}C (30 s) followed by 30 cycles of 95^{o}C (30 s), 55^{o}C (60 s), 68^{o}C (336 s). Reactions were then subject to Dpn I digest for 1 hour at 37^{o}C. 3 μL of this was then transformed into 100 μL of *E. coli* XL10-Gold as per Section 2.3.5. By this process roughly 1000 colonies were achieved following transformation.

2.5.3 Colony picking and library culture

2 mL 96 deep well plates (Fisher Scientific Ltd) were dispensed with 1.5 mL of sterile 2 TY media containing 50 μg mL^{-1} ampicillin using a Genetix plate filler. Colonies from the respective CASTing libraries were picked and used to inoculate six individual wells of the deep well plates. Positive controls (colonies containing the pKK-ItaE construct (Section 2.3.2) were picked and used to inoculate wells H7-9. Negative controls (colonies containing empty pKK-mod (Section 2.3.2) were picked and used to inoculate wells H10-12. A breathable plate seal was used on each plate. Plates were incubated, with shaking (200 rpm) at 30^{o}C for 5 hours. IPTG was then added to a final concentration of 0.1 mM using a 1200 μL electronic multichannel pipette. Cultures were then left to express, with shaking (200 rpm), at 30^{o}C for 16 hours.

2.5.4 Screening fraction preparation

For preparation of the enzyme fraction in Section 3.5.1, 1 g of frozen *E. coli* XL10-Gold pellet containing overexpressed eTA was resuspended in 20 mL of 50 mM sodium phosphate buffer pH 8.0. This was lysed by sonication (Section 2.4.1) and kept on ice for use in the PS cleavage assay.

Cultures from Section 2.5.3 were used to make a glycerol reference plate. 100 μL of culture was added to 100 μL sterile 50% glycerol in 250 μL 96 well plates. These were sealed and stored at -20^{o}C. Cultures were pelleted by centrifugation at 1000 g for 20 minutes and supernatant discarded. Pellets were then frozen at -80^{o}C for 1 hour or until required. 2 mL of degassed lysis buffer was added to each well using a Genetix plate filler and pellets thoroughly re-suspended by pipetting up and down. Plates were left to incubate for 1 hour at 37^{o}C and then centrifuged at 6000 g for 1 hour. The soluble supernatant was then used directly as the fraction for screening.

Lysis buffer
50 mM sodium phosphate, pH 8.0
100 mM NaCl
1 mM EDTA
2 mM $MgCl_2$
0.2 mg mL^{-1} lysozyme
0.05 mg mL^{-1} DNase

2.5.5 A high-throughput screen to identify threonine aldolase stereoselectivity

UV-transparent microwell plates (Greiner Bio-One Ltd) were used for activity measurements. 140 μL of screening buffer containing 50 mM sodium phosphate, pH 8.0 and 50 μM PLP was dispensed into each well using an electronic multiwell pipette (Starstedt, UK). 10 μL of the respective enzyme fraction (Section 2.5.4) was added to the assay buffer using a reverse pipette technique. These were then left to shake (100 rpm)

at 30^{o}C for 10 minutes. 0.6 mM stocks of 2*S*3*R* and 2*S*3*S* phenylserine were made up in 50 mM sodium phosphate buffer, pH 8.0. Screening was performed on a FLUOstar Galaxy plate reader. The pump for automated substrate injection was first cleaned thoroughly by washing with 70% ethanol followed by dH_2O. Several pump volumes of 50 mM sodium phosphate, pH 8.0, were then passed though before the pump was loaded with the necessary PS substrate. Plates were screened using a modified script that involved an initial injection of 50 μL of substrate into all wells with an immediate Abs_{260} reading. Wells were injected and read horizontally (A1-12, B1-12 etc.). This first inject and measure step took 2 min 43 s. The subsequent measurement step involved Abs_{260} readings in the same order as previous at took 2 min 40 s to complete. Four measurement steps were performed in total, values from these were used to determine the activity in each well. An identical procedure was then followed but for screening with the other PS epimer.

2.5.6 Crude lysate further selectivity assay

UV transparent microwell plates (Greiner Bio-One Ltd) were used for further activity measurements. 120 μL of screening buffer containing 50 mM sodium phosphate, pH 8.0 and 58 μM PLP was dispensed into each well using an electronic multiwell pipette (Starstedt, UK). 30 μL of the identified screened fractions was added to the assay buffer using a reverse pipette technique. These were then left to shake (100 rpm) at 30^{o}C for 10 minutes. 0.6 mM stocks of 2*S*3*R* and 2*S*3*S* PS were made. The selectivity screen was performed in triplicate on a FLUOstar Galaxy plate reader. The pump for automated substrate injection was first cleaned thoroughly by washing with 70% ethanol followed by dH_2O. Several pump volumes of 50 mM sodium phosphate, pH 8.0, were then passed though before the pump was loaded with the necessary PS substrate. Individual wells were screened for 1 minute following injections of 50 μL of substrate. Abs_{260} readings were taken every 3 seconds and initial rates calculated. An identical procedure was then followed but for screening with the other PS epimer with all measurements in triplicate.

2.5.7 Purified enzyme selectivity assay

The purified enzyme selectivity assay was performed using the same methods as described in Section 2.5.6 with a few alterations. Rather than adding crude lysate fractions to assay buffer, 50 μg of purified enzyme was added to each well. All buffer components and volumes were adjusted accordingly.

2.5.8 Enzyme kinetics

Steady-state Michaelis-Menten kinetics were performed for the retro-aldol cleavage of 2S3R and 2S3S PS. Fresh stocks of each substrate were made on the same day as the kinetic experiment in 50 mM sodium phosphate, pH 8.0. PLP was also made up as a fresh stock in the same buffer. Enzymes were used as day old preparations following purification (Section 2.4). One hour prior to kinetic analysis, substrates and buffers were incubated in a water bath at 30^o. Enzymes were kept on ice and DMSO was kept at room temperature. Kinetics were measured for a range of substrate concentrations using an appropriate amount of enzyme. Each reaction was performed following a specific order using a clean tip for addition of each reaction ingredient. First 100 μL (10% (v/v)) of DMSO was added to a quartz cuvette (path length 1 cm), 50 mM sodium phosphate was added next followed by the required volumes of substrate. PLP was further added and thorough mixing by pipetting up and down. To start the reaction an appropriate amount of enzyme was added to the cuvette. A lid was placed on top and the cuvette was carefully inverted two times. This was then quickly placed into the Uvicon spectrophotometer for activity measurements. A linear increase in absorbance was followed at 279 nm. A rate for this linear reaction was determined (ΔAbs_{279} min^{-1}) and used to calculate the specific activity of the enzyme. The molar extinction coefficient of 1.4 x 10^{-3} M^{-1} cm^{-1} for benzaldehyde at 279 nm was used as described by Liu et al. [Liu et al., 1998]. Kinetic parameters k_{cat} and K_m were obtained by fitting the data to the Michaelis-Menten equation (equation 2.1) on GraphPad Prism 7.

$$v_o = \frac{V_{max}[S]}{K_m + [S]} \tag{2.1}$$

2.6 Crystallographic methods

The crystallisation conditions for a similar *E. coli* threonine aldolase had previously been described [di Salvo et al., 2014]. Conditions used were optimised around; 0.2 M MgCl2 hexahydrate, 0.1M HEPES (pH 7.5), 30% PEG 400. Crystals were grown by the sitting drop vapour diffusion method and sizeable crystals were observed after 8 days at 20^{o}C.

2.6.1 Data collection and data processing

Data collection was carried out at Diamond Light Source macromolecular crystallography beamline I02. Data were collected from single crystals for all structures at 100 K. The auto processed data from Fast DP, XIA2 3dii, XIA2 3d and DIALS were used to present diffraction data and determine unit cell parameters. The structure for eTA was achieved using molecular replacement in PHASER [McCoy et al., 2007], with a previously solved structure of a homologous protein (PDB file 4LNL).

2.6.2 Refinement

Refinement was performed using REFMAC 5 [Murshudov et al., 1997]. Structures underwent rigid body refinement followed by numerous rounds of restrained refinement. After each round of refinement, model building and adjustments were made using COOT [Emsley and Cowtan, 2004] including the insertion of solvent molecules and metal atoms. The PLP ligand was designed using PRODRG [Schuttelkopf and van Aalten, 2004].

2.7 Computational modelling methods

2.7.1 Definition of non-standard residues

Topology files for the non-standard residues (benzaldehyde and PLP-glycine (PXG)) were generated using the Antechamber package within AmberTools16 and the general AMBER force field. These topology files were used for the subsequent MD simulations.

2.7.2 Molecular mechanics molecular dynamics simulations

Molecular mechanics (MM) molecular dynamics (MD) simulations were performed using the AMBER molecular mechanics ff14SB force field [Maier et al., 2015] and AmberTools16 to perform MD simulations. In all cases 150 psec of MD was simulated using the sander program within AmberTools16. The first 50 psec of simulation were restrained with a force constant of 50 kcal/mol $Å^{-1}$ between the CA atom of benzaldehyde and the CG2 atom of PXG. This also included an initial heating phase to 300K. The restraint was removed for the final 100 psec of simulation.

2.7.3 Quantum mechanics/molecular mechanics (QM/MM) MD simulations

QM/MM MD were performed for umbrella sampling simulations [Torrie and Valleau, 1977]. QM and MM regions were defined as shown in Figure 2.2. The semiempirical QM method, pm6 [Stewart, 2007], was applied to ligands, benzaldehyde and PXG. In PXG the phosphate group and adjoining CH_2 group was excluded from the QM region resulting in a QM charge of -1. QM regions were defined in the input scripts and link atoms placed between QM and MM atoms, where covalently bonded, to fulfil the valences of these atoms. A 20 Å surrounding sphere (about benzaldehyde and PXG) was treated using the AMBER MM ff14SB [Maier et al., 2015] forcefield during simulation. All atoms outside the sphere remained static during simulation. Umbrella sampling involved a -0.1 Å stepwise reduction between the distance of the CA atom of benzaldehyde and the CG2 atom of PXG. For this a force constant of 100 kcal/mol $Å^{-1}$ was applied. A dihedral restraint about the stereochemical coordinate was applied for 3*R* product forming reactions as instances of these "flipping" into a pre-3*S* orientation were observed. For this an upper and lower force constant of 50 kcal/mol $Å^{-1}$ was applied to keep the benzaldehyde ligand between 0-180^o of the stereochemical coordinate (Figure 5.7).

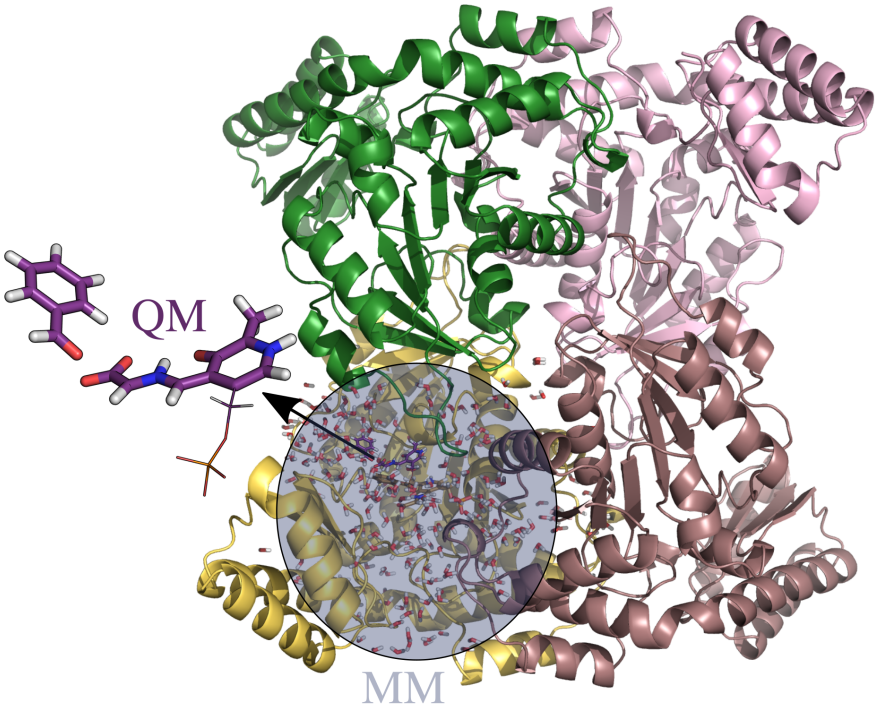


Figure 2.2: QM and MM regions for QM/MM MD umbrella sampling simulations. The total non-static modelling region is contained within the sphere. Everything outside this region remains static. The QM region (purple sticks) contains the atoms belonging to benzaldehyde and PXG ligands (but not the phosphate group and adjoining CH_2). Everything else is in the MM region.

2.7.4 Visualisation of simulations

MM and QM/MD trajectory files were generated using the AmberTools package cpptaj [Roe and Cheatham III, 2013]. These were visualised in PyMOL Molecular Graphics System [DeLano, 2009], Version 1.8.7.0.

Chapter 3

E. coli Threonine Aldolase

3.1 Introduction

Low-specificity L-threonine aldolase (TA) from *Escherichia coli* (eTA) was first described by Liu et al. [Liu et al., 1998]. They found eTA cleaves both diastereomers of L-threonine with around a 42-fold preference for L-*allo*-threonine . While classified as a *threonine* aldolase, eTA demonstrated to be rather more active with L-phenylserine (PS). The stereoselectivity is however reduced with a 1.3-fold preference for the cleavage of 2*S*3*R* PS over 2*S*3*S* PS (Figure 3.1) [Liu et al., 1998]. This poor diastereoselectivity is recurring in L-TAs with substrates other than threonine [Steinreiber et al., 2007, Beaudoin et al., 2018]. PS and PS-derived compounds are considered to be synthetically interesting as intermediates or active ingredients, for example L-DOPS, used in the treatment of Parkinson's [Ogawa et al., 1985] and the antibiotic thiamphenicol [Ratzan et al., 1974] (Figure 3.2).

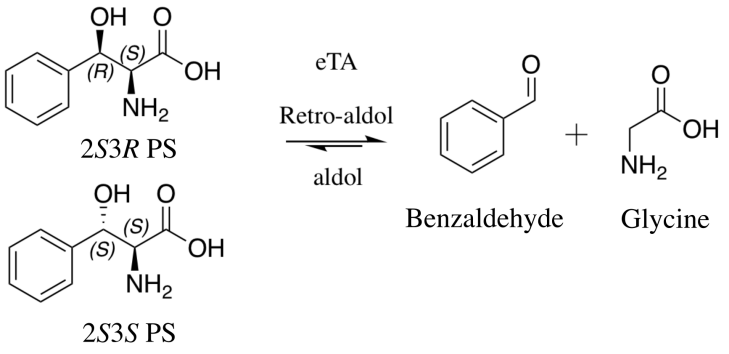
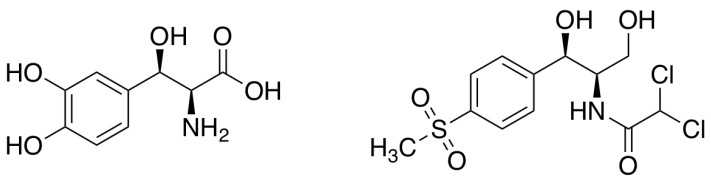


Figure 3.1: The reversible retro-aldol cleavage of 2*S*3*R* and 2*S*3*S* PS into benzaldehyde and glycine. A reaction catalysed by eTA with poor stereoselectivity [Liu et al., 1998].



L-DOPS Thiamphenicol

Figure 3.2: Molecular structure of the Parkinson's drug L-DOPS and the antibiotic thiamphenicol. Both drug compounds display structural similarities to PS.

Synthesis of PS compounds using TAs has focused on exploiting different temperatures, solvents [Chen et al., 2017], phases [Tibhe et al., 2013], and enzyme imobilisation strategies [Tibhe et al., 2013] to obtain improved stereoselectivity. However, few have used enzyme engineering to achieve desirable improvements [Gwon and Baik, 2010]. It would appear this is in part due to lack of an appropriate high-throughput screen. One of the main difficulties is having a suitable method that can discriminate between PS diastereomers under kinetically controlled synthesis. While this can be achieved through high-performance liquid chromatography (HPLC) or nuclear magnetic resonance (NMR), these methods rather limit the throughput of screening. This can be overcome if the direction of the screened reaction is reversed. By following the retro-aldol cleavage of optically pure 2*S*3*R* and 2*S*3*S* PS, two comparable reactions relating to enzyme stereoselectivity are obtained (Figure 3.1). Reisinger et al. [Reisinger et al., 2006] developed a method based on this retro-aldol cleavage of PS. They used a direct colony based assay where in-cell TA cleavage of PS was coupled to a benzaldehyde dehydrogenase. Subsequent oxidation of benzaldehyde into benzoic acid via the dehydrogenase formed NADH which could be measured by fluorescence. Individual colonies harbouring TA variants could be transferred to a membrane and screened for activity with 2*S*3*R* and 2*S*3*S* PS. The success of this method in identifying stereoselective variants was however not reported. [Reisinger et al., 2006].

The aims of this results chapter are to clone the gene for the *low-specificity* L-eTA into a suitable expression system. We shall express, purify and kinetically characterise the enzyme with optically pure 2*S*3*R* and 2*S*3*S* PS. Further we aim to develop a unique, high-throughput screen, applicable for the screening of eTA variants for improved stereoselectivity.

3.2 Cloning the gene for eTA

A synthetic codon optimised gene for the *low-specificity* L-eTA (ItaE) had previously been ordered by the Berry Lab (Appendix Figure A.1). The gene came pre-cloned into a pUC57 vector. This was transformed into XL10-Gold cells (Section 2.3.5). Ligation independent cloning (*fast cloning*) [Li et al., 2011] was used for insertion of ItaE into a pKK223-3 expression vector. Primers were designed to amplify pKK223-3 and ItaE by PCR in a way that allowed the formation of the recombinant expression plasmid (pKK-ItaE) with ItaE present at the multiple cloning site (MCS) (Appendix Table A.1). Clear bands for the plasmid and the gene at the correct size were observed following PCR and agarose gel electrophoresis (Figure 3.3).

Following ligation independent cloning, the pKK-ItaE construct was transformed into XL10-Gold cells (Section 2.3.5). Ten single colonies were selected and cultured overnight in 2 TY media (Section 2.2.4). From these, glycerol reference stocks were stored (Section 2.2.5) and DNA was prepared (Section 2.3.3). Successful cloning of the pKK-ItaE construct was highlighted in restriction digest experiments (Figure 3.3). DNA from lanes 1 and 2 of Figure 3.3C was sent for sequencing analysis using universal primers PTRC-99A-for and PTRC-99A-rev. These would bind either side of the ItaE gene in the MCS as shown in Figure 3.3B. The correct sequence was observed for both DNA samples affirming successful cloning (Figure 3.4).

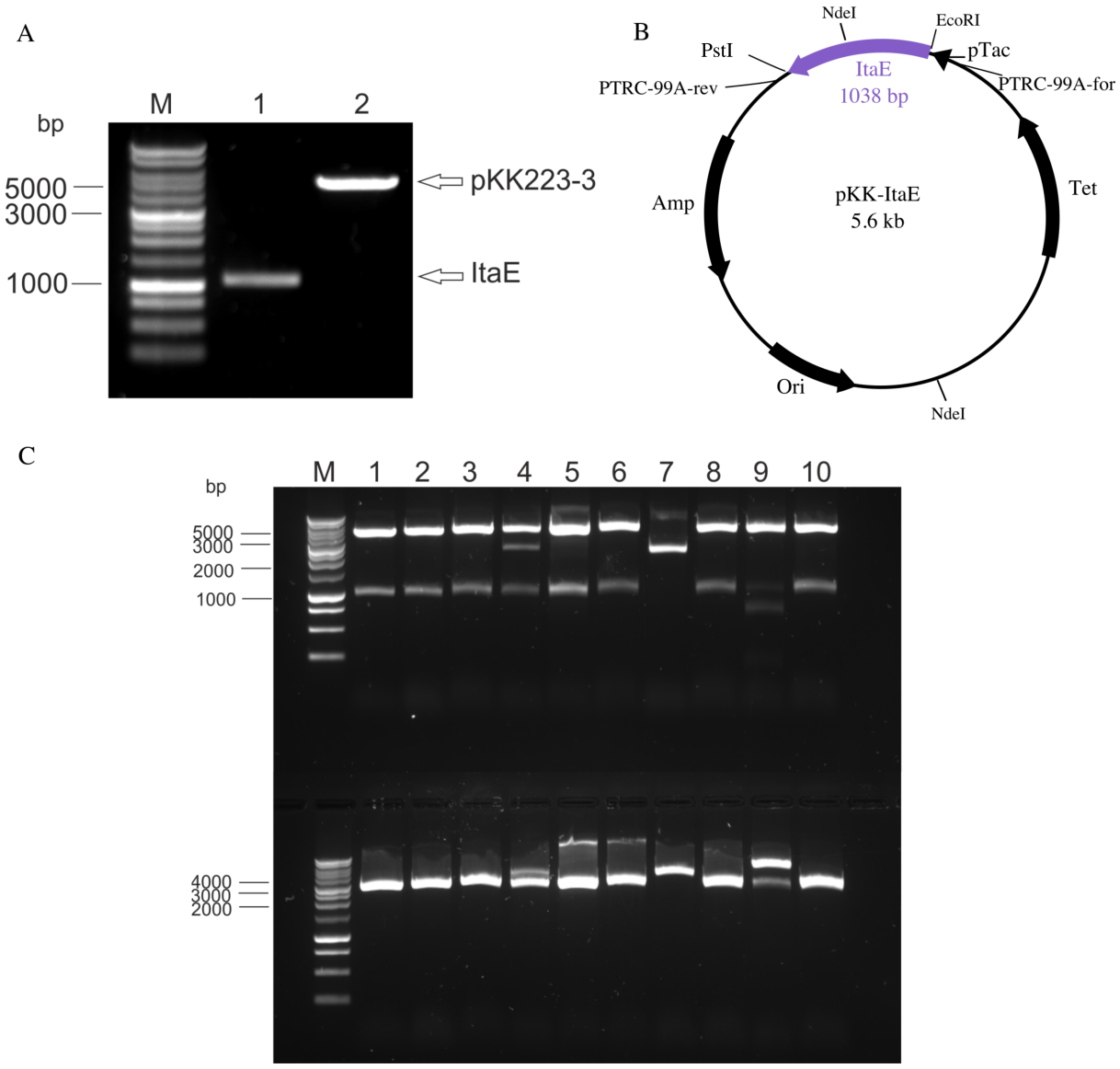


Figure 3.3: A: PCR amplified ItaE and pKK223-3 at the expected sizes. Lane M: 10 kb DNA ladder. Lane 1: PCR amplified ItaE (1 kb). Lane 2: PCR amplified pKK223-3 (4.6 kb). B: The resulting expression plasmid (pKK-ItaE) from ligation independent cloning [Li et al., 2011]. The plasmid contains ampicillin (Amp) and tetracycline (Tet) resistance genes, an origin of replication (Ori) and a tac promoter (pTac). The ItaE gene was cloned into the MCS with flanking PstI and EcoRI restriction sequences. Sites for PTRC-99A-for and PTRC-99A-rev universal primers are highlighted either side of the MCS. C: Agarose gel for the restriction digests of pKK-ItaE constructs. Top: Lane M: 10 kb DNA ladder. Lanes 1-10: Double digested DNA from individual ligation independent cloning transformants. DNA was digested with EcoRI and HindIII restriction enzymes to separate the gene and plasmid to give two bands of 1 kb and 4.6 kb respectively as highlighted in Figure 3.3. Bottom: Lane M: 10 kb DNA ladder. Lanes 1-10: NdeI double digested DNA. pKK-ItaE constructs would be expected to show two bands of 2.8 kb and 2.9 kb. These bands would not be expected to be visible individually, rather a single band between the 2.5 kb and 3 kb ladder markers.

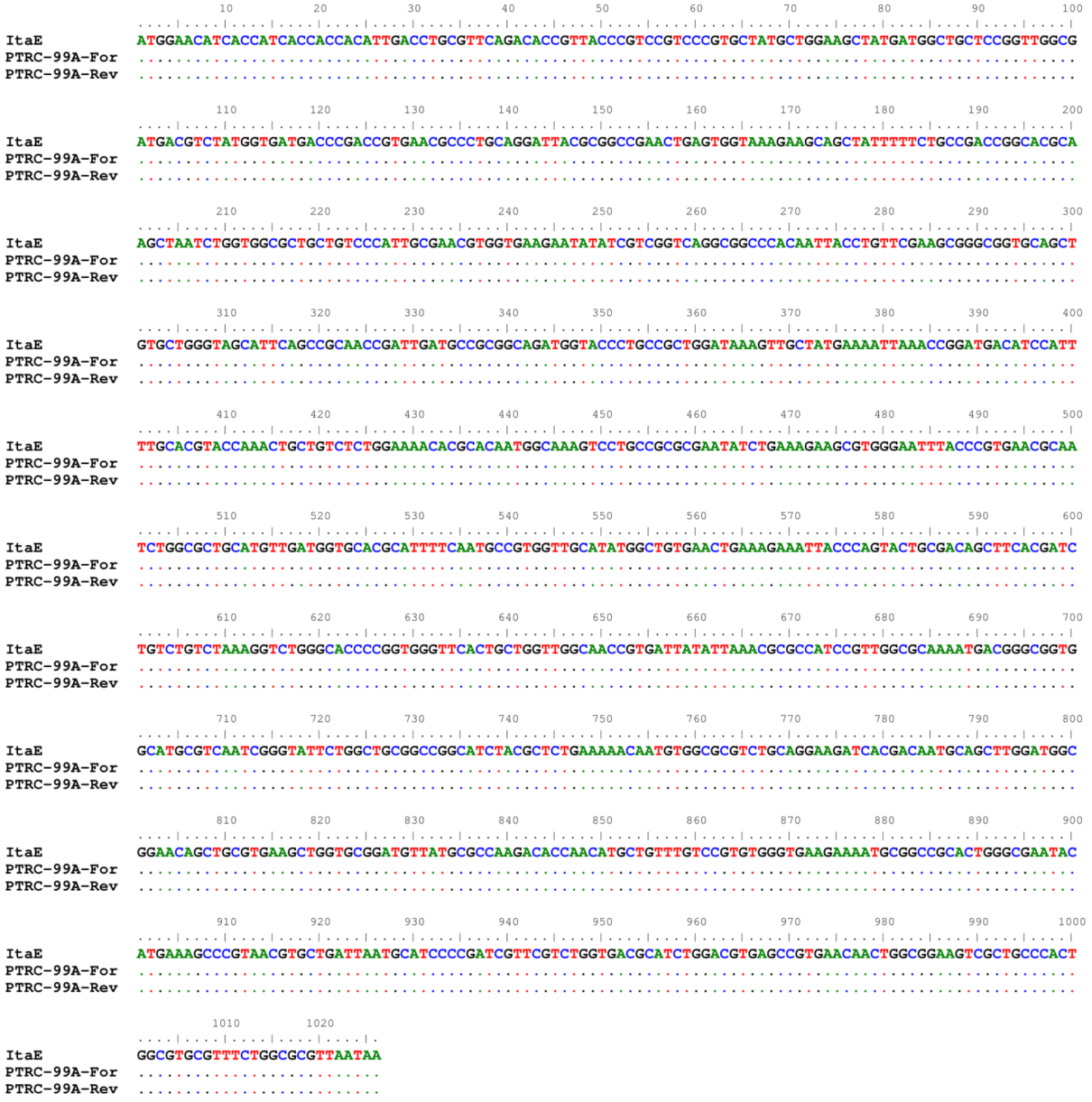


Figure 3.4: Sequencing data for correct construct obtained from ligation independent cloning. Results were aligned to the codon optimised ItaE gene sequence. For the reverse alignment the sequence data was reverse complimented so pairwise data could be observed.

3.3 Expression and purification of eTA

The tac promoter in pKK-ItaE (Figure 3.3B), is a hybrid of the trp and lac promoters [De Boer et al., 1983]. Expression of genes under this promoter can occur by induction with IPTG. Several aldolase enzymes used in the lab were shown to express at high levels in 2 TY media following IPTG induction. A similar protocol was therefore followed (Section 2.2.4). pKK-ItaE was transformed into *E. coli* BL21 cells and left to grow on 2 TY agar containing

50 μg mL^{-1} ampicillin. Single colonies were picked and used to inoculate 5 mL overnight cultures of 2 TY supplemented with 50 μg mL^{-1} ampicillin. Glycerol stocks of *E. coli* BL21-pKK-ItaE cultures were stored and a further 50 μL of the culture was used to inoculate a 50 mL day culture. This was incubated for a further 8 hours at 37^{o}C, with shaking. 10 mL of day culture was used to inoculate 1 L of 2 TY supplemented with 50 μg mL^{-1} ampicillin. Cultures were incubated at 37^{o}C, with shaking until an optical density (O.D) of 0.6 at 600 nm was reached. Protein expression was induced by addition of IPTG to a final concentration of 0.1 mM.

eTA, encoded by the ItaE gene, was purified using a batch method by nickel affinity chromatography (Section 2.4.1). The His_6-tag on eTA enabled efficient isolation of the enzyme. Samples were taken throughout the purification for analysis by SDS-PAGE (Figure 3.5). An overexpressed band was observed around 37 kDa in the cell lysate fraction. The same band was also observed at a similar level in the soluble and unbound fractions suggesting the nickel resin to have been overloaded. The eluted fraction showed a single band for purified eTA corresponding to the expected mass of the enzyme (37 kDa). eTA was dialysed into 50 mM sodium phosphate buffer (pH 8.0) and an aliquot was prepared and sent for accurate mass determination by mass spectrophotometry (Figure 3.6). The peak in the deconvoluted spectrum was within 1 Da of the expected mass of eTA and is within the error limit of the instrument. This provided final conformation that eTA had been successfully cloned, expressed and purified and was ready for further enzymatic characterisation.

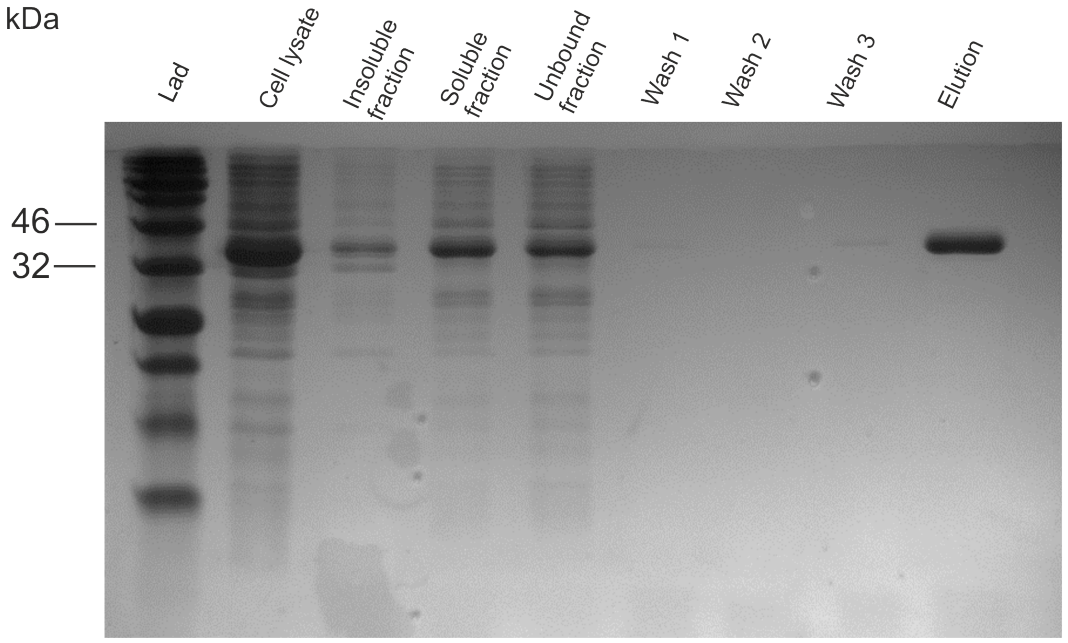


Figure 3.5: SDS-PAGE gel for the purification of eTA. Samples were loaded as labelled above and ran at a 20-fold dilution. Large amounts of overexpressed eTA remain in the unbound fraction suggesting the resin to be overloaded with the recombinant protein. The eluted fraction shows a very strong single band at the expected size for eTA (37 kDa) with no suggestion of contaminating proteins.

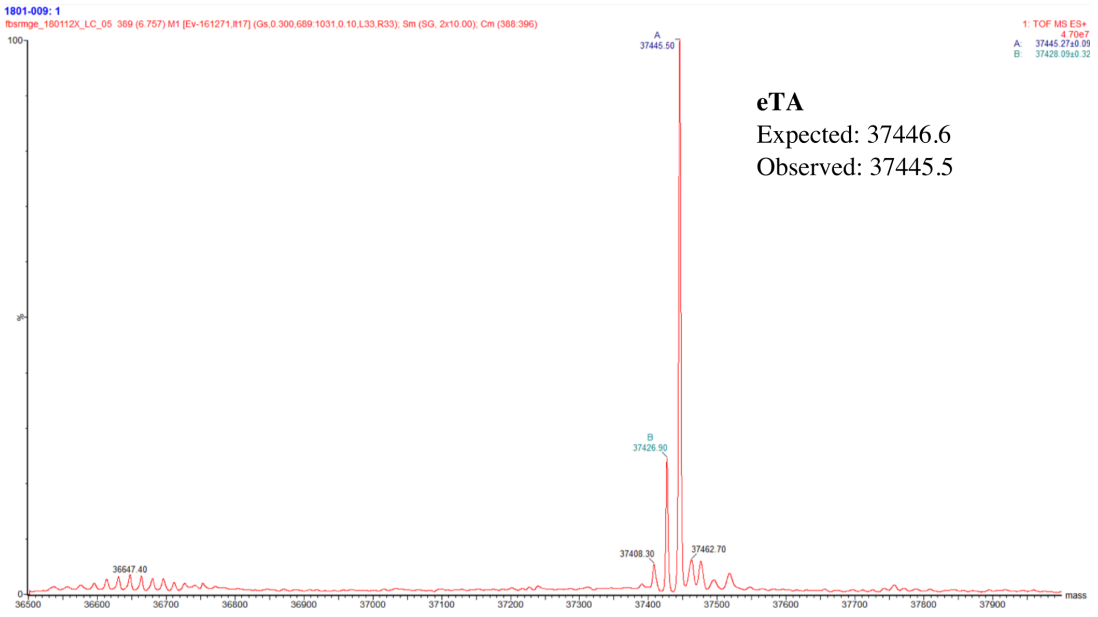


Figure 3.6: Deconvoluted mass spectrum for eTA. The main peak shows a mass of 37446 Da which is within 1 Da of the expected mass of eTA (37447 Da). Sample was analysed under denaturing conditions in acetonitrie.

3.4 Characterisation of eTA stereoselectivity with 2*S*3*R* and 2*S*3*S* PS

eTA has been shown to be poorly stereoselective in the retro-aldol cleavage of 2*S*3*R* and 2*S*3*S* PS [Liu et al., 1998]. This provides an ideal system (enzyme and substrates) to target for improving the stereoselectivity. 2*S*3*R* and 2*S*3*S* PS substrates were ordered as single enantiomer compounds (Section 2.2).

3.4.1 An activity assay for the reversible cleavage of PS

Benzaldehyde and PS display different UV absorbance spectra due to the different conjugation of electrons within their aromatic rings. To observe this the spectra of benzaldehyde and 2*S*3*R* PS were measured. 1 mM of each compound was added to 50 mM sodium phosphate buffer, pH 8.0. Separate measurements of each compound were performed and the spectra recorded (Figure 3.7). A distinct λ_{max} at 250 nm was observed for benzaldehyde that was not apparent for 2*S*3*R* PS. This difference in absorbance between substrate and product spectra allows the aldol and retro-aldol reactions in Figure 3.1 to be followed continuously. To demonstrate this two separate reactions were performed. In the first 50 μM eTA was added to 1 mM 2*S*3*R* PS and the spectra recorded over 5 minutes (Figure 3.8A). The results show clear production of benzaldehyde from the characteristic peak at 250 nm. Interestingly, the results also suggest that any wavelength between 230-300 nm is suitable for following the retro-aldol cleavage of PS.

In the second reaction 200 μg of eTA was added to 1 mM benzaldehyde, 30 mM glycine and the spectra recorded. It is necessary to use an excess of glycine to shift the unfavorable aldol equilibrium [Franz and Stewart, 2014] and observe PS formation (Figure 3.8B). A reduction in the benzaldehyde peak suggests it has been consumed in the aldol condensation reaction however this provides no information on the stereoselectivity of the reaction.

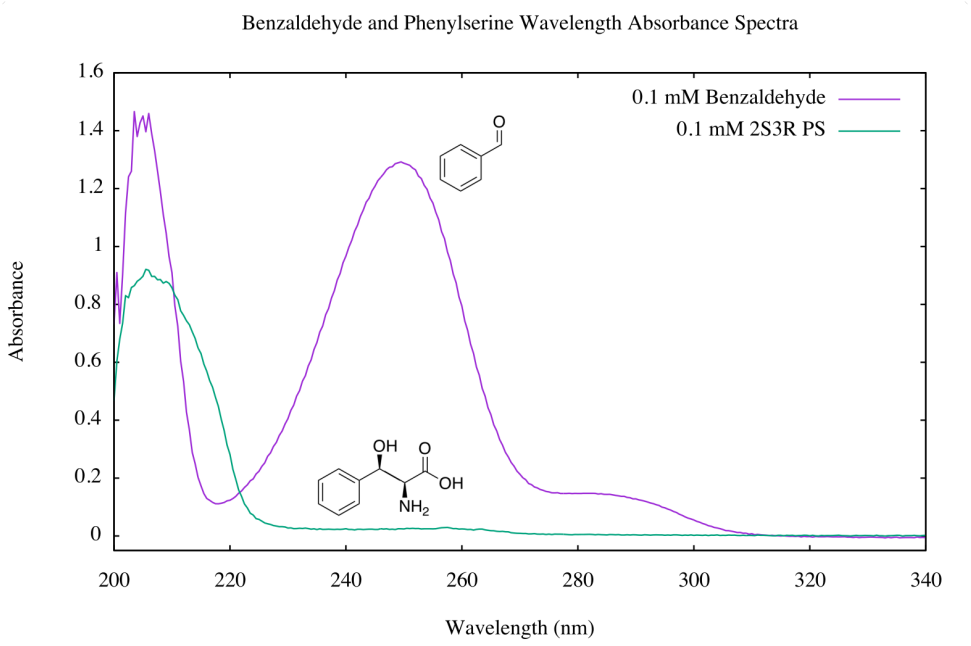


Figure 3.7: Wavelength absorbance spectra for benzaldehyde and 2*S*3*R* PS measured using a JASCO spectrophotometer with a path length of 1 cm. A solution containing 0.1 mM benzaldehyde (purple) in 50 mM sodium phosphate pH 8.0 was measured against a reference cuvette (pathlength 1 cm) containing the same buffer. The spectrum for 2*S*3*R* PS (green) was also measured at 0.1 mM concentration in 50 mM sodium phosphate buffer, pH 8.0.

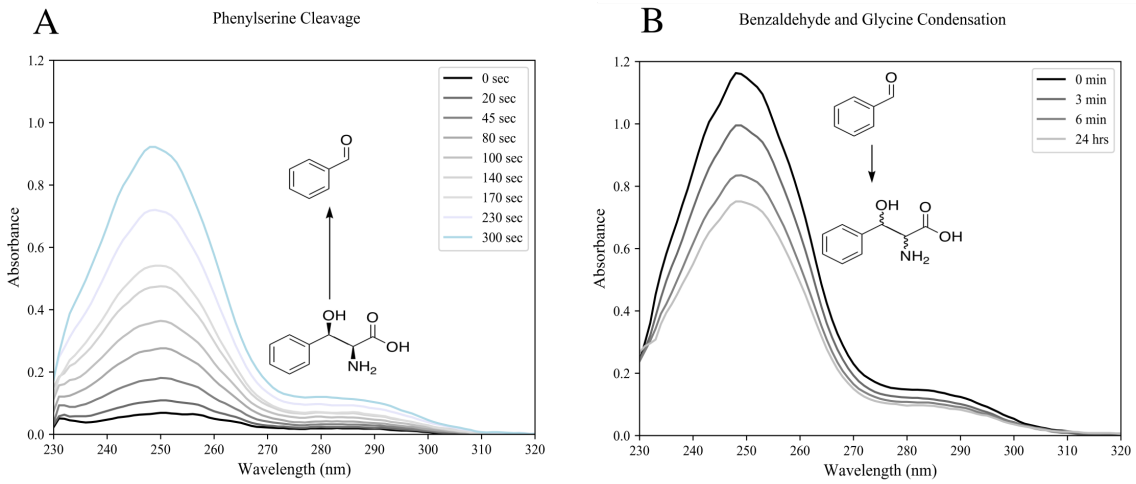


Figure 3.8: A: Cleavage reaction progression. 50 μg of eTA was added to 1 mM 2*S*3*R* PS in 50 mM sodium phosphate buffer pH 8.0 and 50 μM PLP in a 1 mL reaction volume. Spectra were recorded over 5 minutes using a Nanodrop 2000 with 1 mm pathlength. B: Condensation reaction progression. 200 μg of eTA was added to 1 mM benzaldehyde, 30 mM glycine in 50 mM sodium phosphate buffer pH 8.0 and 50 μ M PLP in a 1 mL reaction volume. Spectra are shown at 3 minute intervals and a final spectrum measured after 24 hours.

3.4.2 eTA kinetic characterisation

Michaelis-Menten kinetics were performed on purified eTA at a range of 2*S*3*R* and 2*S*3*S* PS concentrations (Section 2.5.8). The specific activity of eTA at each substrate concentration was calculated using the molar extinction coefficient of 1.4 x 10^{-3} M^{-1} cm^{-1} for benzaldehyde at 279 nm [Liu et al., 1998]. These values were plotted against substrate concentration and Michaelis-Menten curves produced (Figure 3.9) giving kinetic parameters K_m and k_{cat} (Table 3.1). The results show virtually identical results for the cleavage of both L- epimers with specificity constants (k_{cat}/K_m) of 5750 $min^{-1}mM^{-1}$ for the 2*S* 3*R* and 5650 $min^{-1}mM^{-1}$ for 2*S* 3*S* PS. The selectivity preference (SP) compares these two values for a measure of stereoselectivity as given by equation 3.1.

$$SP = \frac{(k_{cat}/K_m)_{3S}}{(k_{cat}/K_m)_{3R}} \tag{3.1}$$

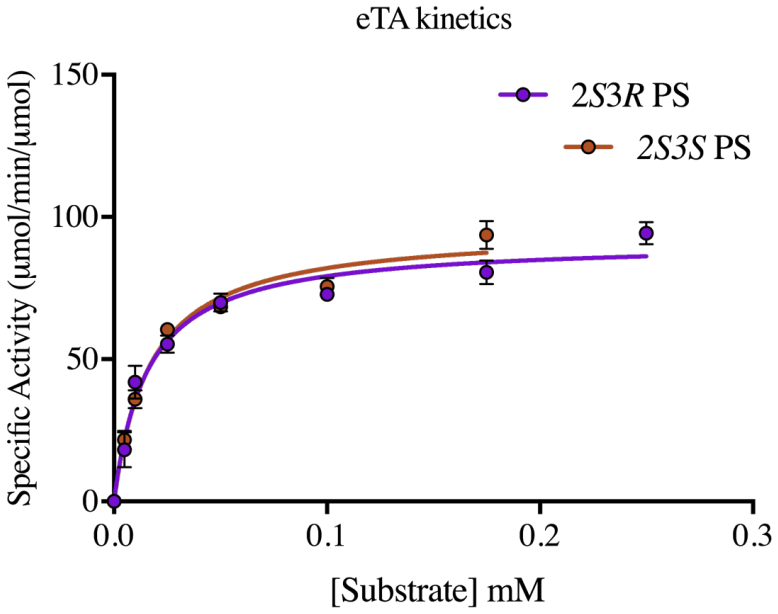


Figure 3.9: Michaelis-Menten kinetics curve for cleavage of both 2*S*3*R* and 2*S*3*S* PS epimers. Measurements were performed in 50 mM sodium phosphate buffer pH 8.0, 50μM PLP, 10% (v/v) DMSO at 30 ^{o}C. Each measurement was performed in a final volume of 1 mL in a quartz cuvette with a 1 cm path length . Reactions were started by addition of an appropriate amount of enzyme and initial rates measured.

	2*S*3*R* Phenylserine			**2*S*3*S* Phenylserine**			
Protein	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	SP 3* (*S*/*R*)
wild-type eTA	92 ± 3	0.016 ± 0.002	5750 ± 740	96 ± 3	0.017 ± 0.002	5650 ± 690	1.0

Table 3.1: Kinetic parameters for the eTA catalysed cleavage of 2*S*3*R* and 2*S*3*S* PS. The selectivity preference (SP) at the 3 position is calculated from k_{cat}/K_M ratios for both 2*S*3*R* and 2*S*3*S* PS

3.5 Development of a high-throughput assay for stereoselective screening.

If improvements in stereoselectivity are to be engineered then a suitable screen is required. It is also necessary to consider the thermodynamics with tasks of this nature (Figure 3.10). Under kinetically controlled conditions the epimerisation barrier ($\Delta\Delta G_{3R}$ / $\Delta\Delta G_{3S}$) will determine the stereoselectivity where the transition state energies (E•PS‡) are representative of the the k_{cat}/K_m value of the reaction. As this energy barrier difference will remain the same independent of the aldol or retro-aldol reaction, we would expect stereoselective variants to be equally selective in both synthesis and cleavage reactions. The retro-aldol reaction will therefore be used as both optically pure 2*S*3*R* and 2*S*3*R* PS compounds are owned. Furthermore we are able to make use of the favoured position of equilibrium.

The screen must be suitable for the throughput required. Given that the structure of eTA is available, this provides a very good starting point for a rational approach where key residues can be identified. CASTing, employed by Reetz et al. [Reetz et al., 2006](Section 1.2.2.1), uses an effective strategy to sample a hotspot within a mutagenesis landscape. It was decided that CASTing would be used for engineering efforts as multiple stereoselective enzymes have been created with this approach [Sandström et al., 2012, Rowles et al., 2016, Prasad et al., 2011]. The calculations for adequate screening of CASTing libraries are discussed in Section 1.2.2.1. However, from previous studies, screening of around 3000 variants was suggested [Reetz et al., 2006, Prasad et al., 2011]. With this, a plate based screen seemed the most appropriate way to proceed. In Section 3.4.1 we had shown an effective spectrophotometric method to monitor the cleavage of PS using purified eTA. If the

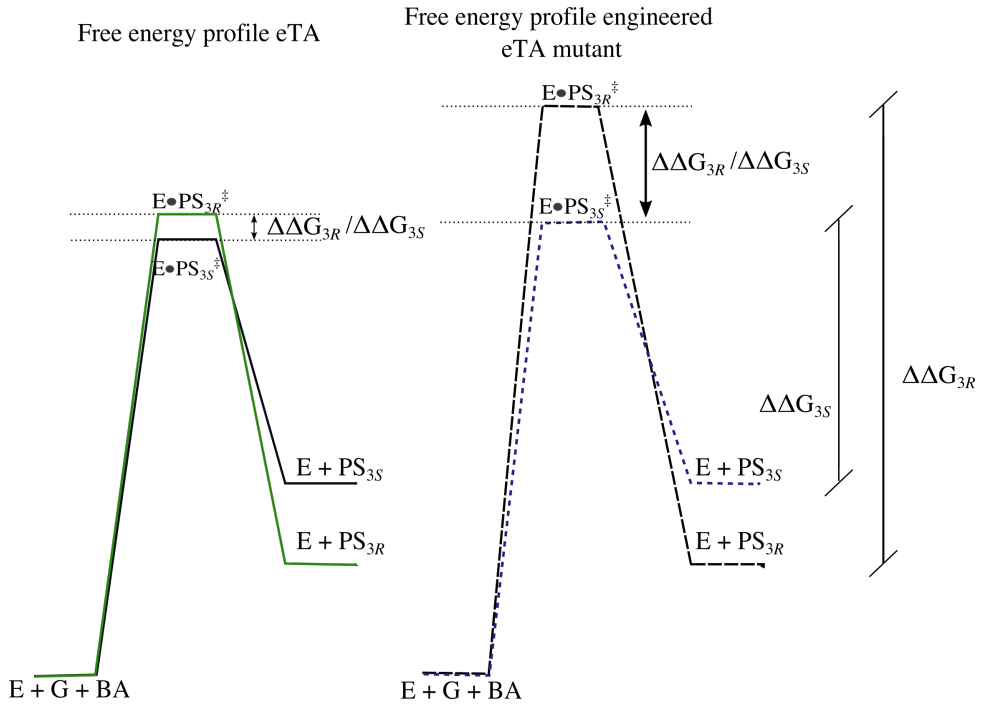


Figure 3.10: A hypothetical free energy diagram showing the reaction for a simplified eTA (left) catalysed PS condensation/cleavage where E is enzyme, G is glycine, BA is Benzaldehyde and PS is phenylserine. The respective 2*S*3*R* and 2*S*3*S* diastereomers are denoted by 3*R* and 3*S* respectively. The left profile shows the wild-type eTA catalysed reaction. In all instances the Gibbs free energy of the starting products, free benzaldehyde and glycine, will remain constant. The Gibbs free energy for both 2*S*3*R* and 2*S*3*S* PS epimers will also remain fixed and the difference between these two energies will determine the thermodynamic ratio once equilibrium is reached. Under kinetic conditions the ratio of the epimers will be dictated by the difference in the energy barriers $\Delta\Delta G_{3R}/\Delta\Delta G_{3S}$. The right profile shows a hypothetical free energy profile for an improved 2*S*3*S* PS stereoselective eTA variant.

same conversion could be shown using crude lysate this will provided a sufficient method for monitoring eTA activity.

3.5.1 Crude lysate activity assay

As shown in figure 3.8 the λ_{max} for benzaldehyde is around 250 nm. Detecting changes in absorbance around this region can be difficult due to large amounts of background absorbance. Several experiments were set up to investigate the cleavage of 2*S*3*R* PS using an eTA crude lysate fraction (Section 2.5.4). Figure 3.11A showed clean spectra could be obtained that followed the cleavage reaction progression when using a reference cuvette. Removing the reference cuvette showed the change in spectra could be observed sitting on top of the large background absorbance peak from the crude soluble extract (Figure 3.11B).

To ensure the reaction being observed was eTA mediated, a negative control was performed using a cell extract containing empty pKK-mod (Figure 3.11C). From these results it was clear the reaction could be observed using a eTA crude lysate at the given conditions.

It should be noted that for the experiments in this section the eTA cell lysate was prepared from pellets produced by large scale expression (Section 2.2.4). This method reliably produced large amounts of recombinant eTA whereas the screening method would require suitable levels of expression in a 2 mL culture. To see if activity could still be observed at this scale a plate based assay was developed.

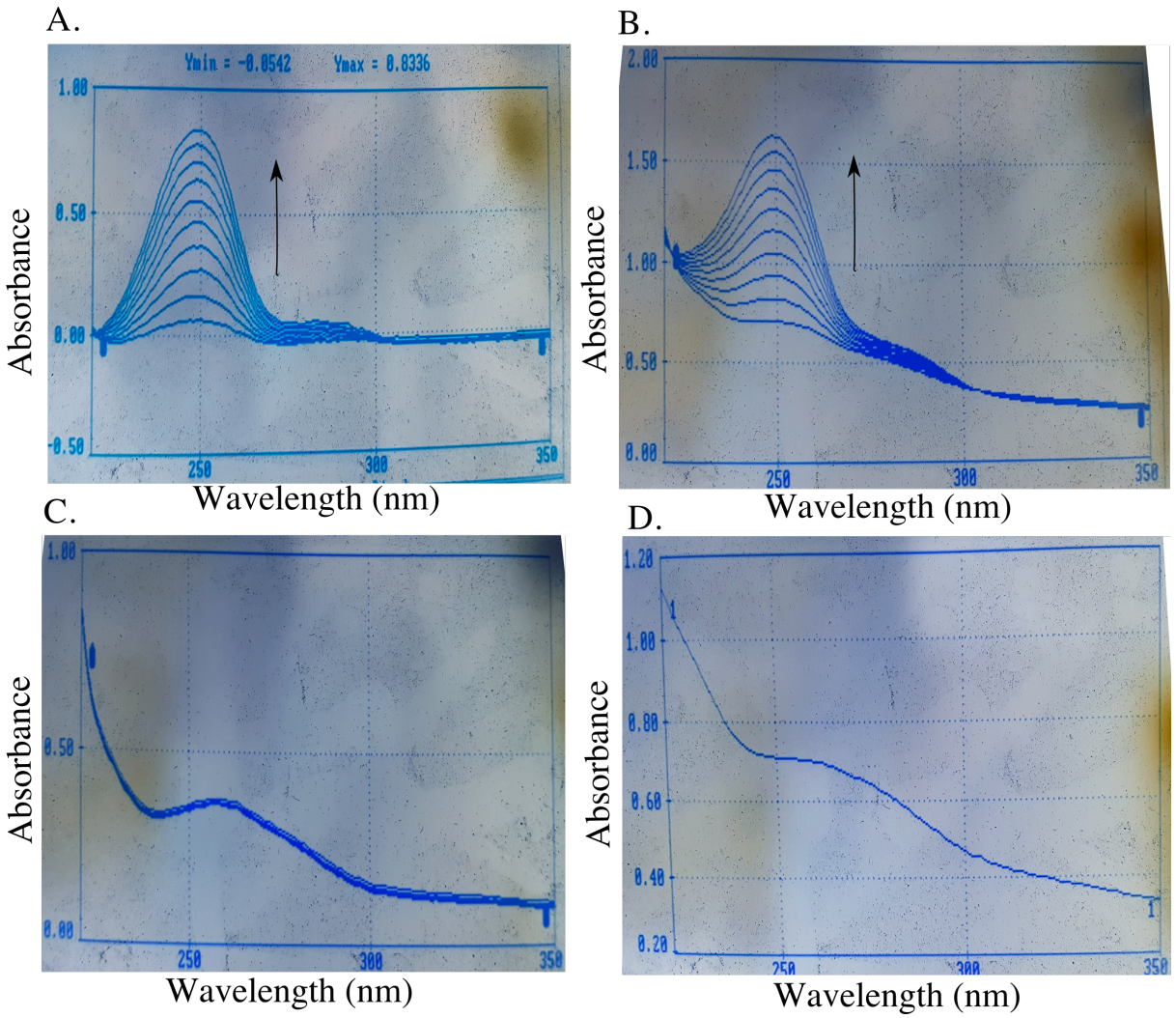


Figure 3.11: Wavelength absorbance spectra for crude cell fractions from a Uvicon spectrophotometer. A: Activity for 5 μL of crude lysate fraction was measured using a reference to hide background absorbance. The enzyme containing fraction was prepared as per the methods in Section 2.5.4 and added to to a reaction buffer containing 50 mM sodium phosphate pH 8.0, 50 μM PLP and 0.1 mM 2*S*3*R* PS in a total reaction volume of 1 mL. Spectra were recorded every minute over a 10 minute period. B: The reference cuvette was removed and the reaction was repeated. This showed the background absorbance underneath the peak being measured. C: Negative control showing a similarly prepared fraction as in A, containing *E. coli* XL10-Gold pKK-mod crude lysate. The fraction was assayed by the same methods as in A. D: Absorbance spectrum for 5 μL of the crude lysate fraction in 1 mL sodium phosphate buffer pH 8.0. This amount appears to be suitable for observing eTA activity without saturating the detector in the spectrophotometer.

3.5.2 Small-scale expression test

Three different controls containing empty *E. coli* XL10-Gold, XL10-Gold ItaE and XL10-Gold pKK-mod (containing only the pKK223-3 plasmid) were grown from single colonies in 2 mL deep well plates (Section 2.5.3). Following IPTG induction and overnight expression, samples were lysed, pelleted (Section 2.5.4) and each crude lystate soluble fraction analysed by SDS PAGE (Figure 3.12). Those fractions containing the recombinant pKK-ItaE showed a clear overexpressed band at the correct size for eTA (37 kDa).

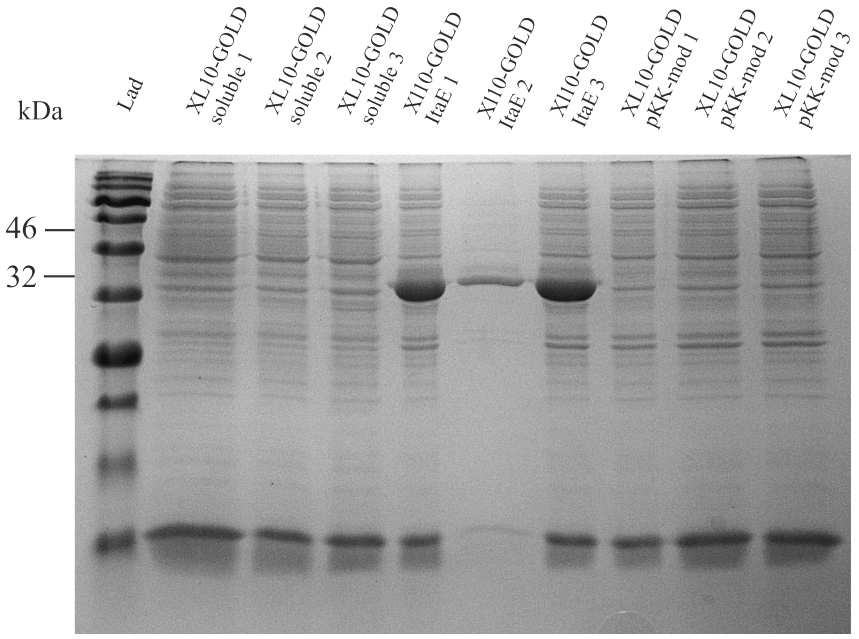


Figure 3.12: SDS PAGE gel showing soluble fractions for each small scale expression test. Wells contained the two negative control fractions: *E. coli* XL10-Gold soluble and XL10-Gold pKK-mod soluble and the positive control fraction XL10-Gold ItaE containing the expressed eTA. For the positive control fractions a strong overexpressed band is observed at the expected size for eTA (37 kDa).

3.5.2.1 Small-scale activity assay

To ensure sufficient levels of activity in the eTA containing fraction an activity assay was performed. 50 μL of the crude lysate fractions were added to 0.15 mM of 2*S*3*R* or 2*S*3*S* PS. The rates of reactions were measured and plotted for each epimer (Figure 3.13). Almost identical levels of activity were observed for the fractions containing overexpressed eTA, whereas the two negative control fractions showed no activity for either of the epmiers.

Both XL10-Gold ItaE and XL10-Gold pKK-mod fractions were used to visualise activity in a plate assay.

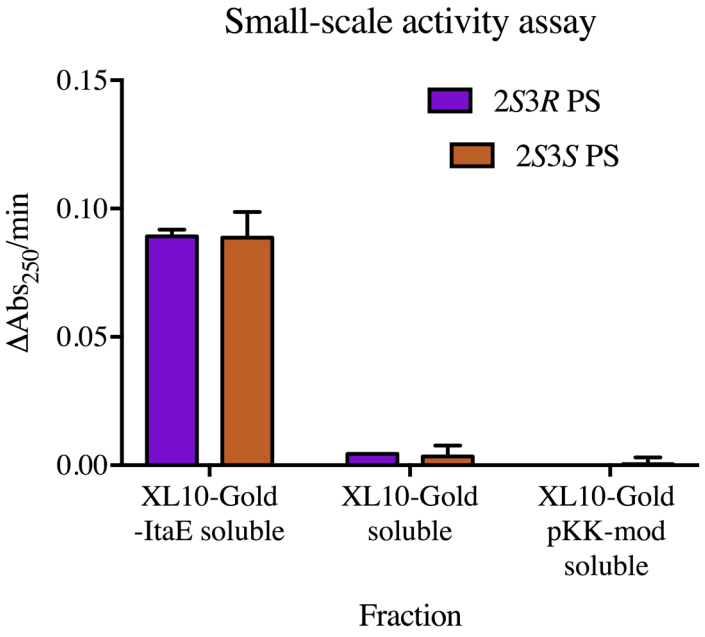


Figure 3.13: Initial rates for the cleavage of 2S3R and 2S3S PS with positive and negative control fractions. Rates were measured using the activity assay from Section 3.4.1 and rates calculated from the change in absorbance at 250 nm. 50 μL of each soluble fraction were added to 50 mM sodium phospahte buffer pH 8.0 containing 50 μM PLP. Each reaction contained 0.15 mM of the respective PS diastereomer. Activities were measured on a Nanodrop 2000.

3.5.3 Plate screen development

To be effectively applied in a high-throughput screen, the crude lysate activities observed in previous sections must be observable in a 96-well plate. Conditions to satisfy this requirement were identified as described. eTA containing crude lysate soluble fractions were prepared as per the methods in Section 2.5.4. To determine a suitable amount of lysate to use in the plate screen, different volumes were assayed and their 2*S*3*R* PS retro-aldol activity observed over 10 minutes (Figure 3.14A). It was suggested that 10 μL of the crude lysate soluble fraction was the optimum amount. This would keep absorbance within the detection range of the plate reader with significant observable activity. A repeat assay of positive and negative control fractions was performed at the identified conditions (Figure 3.14B). It was noticed that the error at the beginning of the reactions was larger

than expected. This was due to an unexpected decrease in absorbance in one of the measurements over the first 50 s (Figure 3.14B). Automated injection of substrate by the plate reader in some instances causes the formation of bubbles. This altered the path length of the well and the absorbance. Degassing buffers, thorough cleaning of injector apparatus and varying concentrations of DNase and lysozyme in the lysis buffer failed to completely eliminate the problem. As a result it was decided that the initial absorbance reading at time=0 would not be used when measuring the rate. A further solution to eliminating the chance of false positives was to perform a secondary activity assay on hits with triplicate measurements.

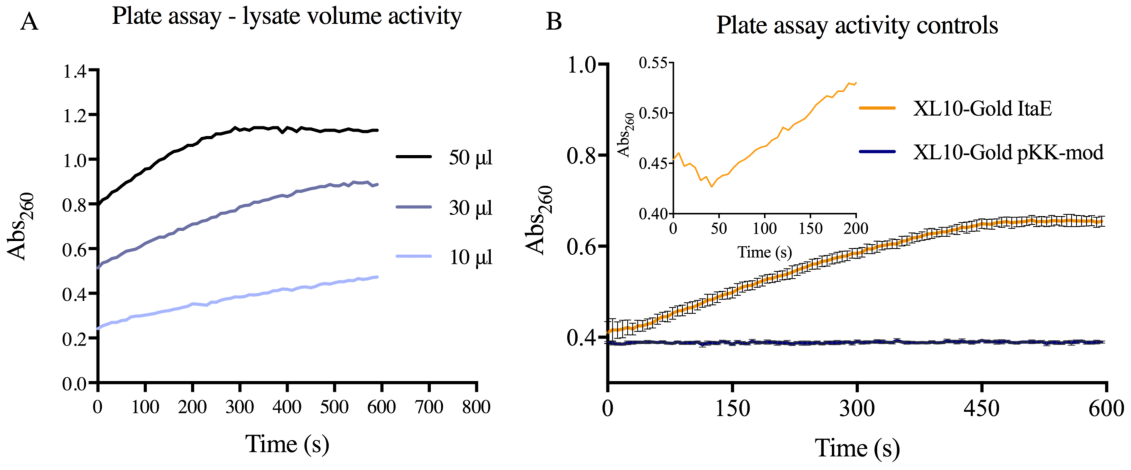


Figure 3.14: A: Different volumes of the eTA crude lysate soluble fractions were tested to find a suitable range for the plate screen. 10 μL gives sufficient activity while remaining within the linear absorbance detection range of the plate reader. 30 μL and 50 μL assays show very clear conversion however the reactions appear to reach equilibrium before 10 minutes. Each reaction contained the given volume of soluble fraction, 50 mM sodium phosphate buffer pH 8.0 and 50 μM PLP. Reactions were started by automated injection of 2*S*3*R* PS to a final concentration of 0.15 mM. Measurements were taken every 6 seconds for 10 minutes using a FLUOstar Galaxy plate reader. B: Continuous activity following the cleavage of 2*S*3*R* PS by increase in absorbance at 260 nm in the presence of 10 μL of soluble XL10-Gold ItaE fraction. Negative control reactions containing XL10-Gold cells with empty pKK-mod showed no increase in absorbance suggesting no PS cleavage. Reactions contained 50 mM sodium phosphate buffer pH 8.0, 50 μM PLP and 10 μL of either the positive or negative control fraction. Reactions were performed in triplicate and started by automated injection of 2*S*3*R* PS to a final concentration of 0.15 mM. Measurements taken every 6 seconds for 10 minutes using a FLUOstar Galaxy plate reader. One positive control shows an initial decrease in absorbance over the first 50 seconds due to formation of bubbles following substrate injection.

The minimum required concentration of PS to observe activity is around 0.15 mM. The screen will therefore only identify stereoselective variants with a K_m greater than this value.

The initial rates measured for PS epimers in the assay will be roughly proportional to k_{cat}/K_m which will give an idea of stereoselectivity as discussed in Section 3.5.

3.5.4 Validity of the screen

The full process of the developed screen, from library to stereoselective mutant, is shown in Figure 3.15. With the screen working for positive and negative controls a blind screening assay was performed. A 2 mL deep well plate was inoculated with positive and negative control single colonies and note made of which went in each. The screening methods in Section 2.5.5 were followed and activities of the wells measured (Figure 3.16). Impressively the screen worked in identifying all of the positive control wells where a threshold change in absorbance of of 0.2 was set (ΔAbs_{260}). One single well identified a positive control that was not expected. It is thought this is likely due to a cross contamination during inoculation. The results suggest the screen is effective in identifying eTA activity. In the following Chapter this screen will be implemented for identifying improved stereoselective variants for PS by CASTing.

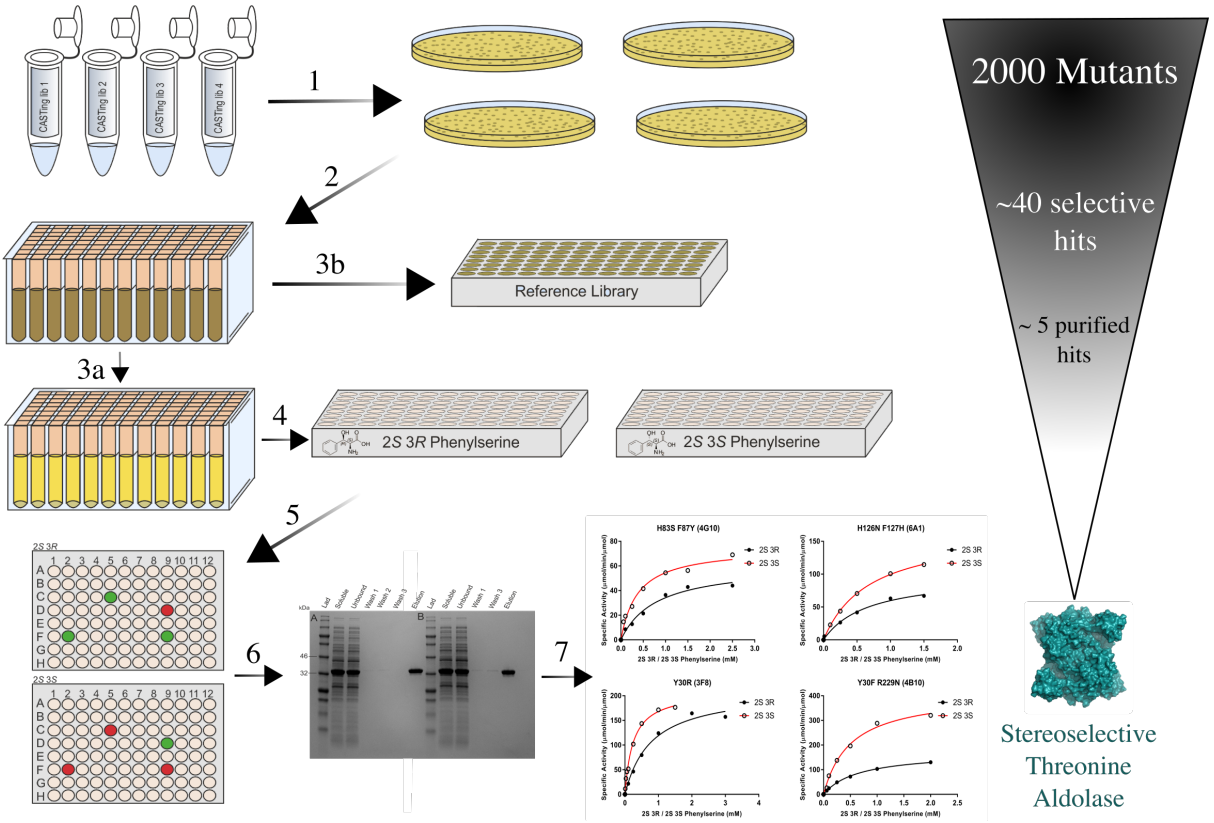


Figure 3.15: A schematic diagram of the developed screen with suitable throughput for screening CASTing libraries. 1: CASTing libraries are produced by PCR using degenerate NDT codons and transformed into XL10-Gold supercompetent cells. 2: Colonies are picked and grown in 2 mL deep well plates before expression is induced by addition of IPTG. 3a: Cells are pelleted by centrifugation and re-suspended in a lysis buffer. Cells debris is next pelleted to obtain a crude lysate soluble fraction 3b: Reference libraries are made from cell fractions and stored in 25% (v/v) glycerol at -80^{o}C. 4: PS screening plates are set up. Plates only differ in the PS epimer used (2*S*3*R* or 2*S*3*S* PS). The exact same amount of enzyme containing fraction is added to the plates before the reaction is started by addition of the substrate by a plate sample injector. 5: Activities are analysed by change in Absorbance at 260 nm. Hits with different observable activities for a single epimer are further assayed to eliminate the chance of false positive results. 6: Selective hits are purified using nickel affinity chromatography. 7: Purified enzymes are kinetically characterised for quantitative assessment of increase in stereoselectivity.

Map	1	2	3	4	5	6	7	8	9	10	11	12
A	1	2	3	4	5	6	7	8	9	10	11	
B	12	13	14	15	16	17	18	19	20	21	22	
C	23	24	25	26	27	28	29	30	31	32	33	
D	34	35	36	37	38	39	40	41	42	43	44	
E	45	46	47	48	49	50	51	52	53	54	55	
F	56	57	58	59	60	61	62	63	64	65	66	
G	67	68	69	70	71	72	73	74	75	76	77	
H	78	79	80	81	82	83	84	85	86	87	88	

Plate	1	2	3	4	5	6	7	8	9	10	11	12
A	-0.04	-0.08	0.00	-0.04	-0.05	0.30	-0.01	0.03	-0.05	-0.03	-0.04	
B	-0.03	0.01	0.58	-0.01	-0.04	-0.03	-0.04	0.03	0.09	-0.03	-0.07	
C	-0.17	0.07	-0.05	-0.01	0.00	0.09	-0.02	0.27	0.50	0.05	0.15	
D	-0.02	0.05	0.04	0.19	-0.10	0.45	0.50	0.10	-0.09	-0.30	-0.06	
E	0.10	-0.03	0.05	-0.05	-0.04	0.46	0.35	-0.08	-0.06	-0.03	0.04	
F	-0.32	0.02	-0.02	0.00	0.18	-0.10	-0.10	-0.04	-0.04	0.63	-0.16	
G	-0.04	0.57	0.00	0.61	-0.01	-0.01	-0.01	0.59	-0.01	-0.03	-0.05	
H	-0.07	-0.03	-0.02	-0.03	-0.02	0.00	-0.04	-0.04	-0.07	0.07	-0.03	

Figure 3.16: Blind screen to identify positive and negative control wells. A 2 mL deep well plate containing growth media was inoculated with single positive and negative control colonies. Positive colonies were added to the plate as shown by yellow boxes in the map plate. The activity assay was performed measuring the 2*S*3*R* PS cleavage. All positive colonies were identified. A threshold activity of 0.2 Δ Abs_{260} for the screen was set. One additional well (G4) was identified as a positive which was possibly due to cross contaminated during inoculation. The activity screening plate contained 50 mM sodium phosphate pH 8.0, 50 μM PLP, and 10 μL of the soluble enzyme fractions. The reactions were started by addition of 2*S*3*R* PS to a final concentration of 0.2 mM

3.6 Summary

The gene for the *low-specificity* L-threonine aldolase from *E. coli* has successfully been cloned using a ligation independent approach. This was successfully expressed, purified and kinetically characterised for the cleavage of both 2*S*3*R* and 2*S*3*S* PS. The eTA has shown to have no stereoselectivity preference for either of the L-PS epimers (1.02 2*S*3*R*), providing a perfect system for engineering stereoselectivity.

An assay for the retro-aldol cleavage of PS has been developed showing any absorbance between 230-300 nm can be used to follow activity. This has further been applied in the development of a high-throughput screen capable of measuring the activity of eTA in crude lysate. With this screen in place we now have a method of detecting stereoselective eTA variants if paired with an effective mutagenesis strategy.

Chapter 4

Directed Evolution of *E. coli* Threonine Aldolase

4.1 CASTing, a method for directed evolution

Many previous enzyme evolution studies have utilised methods such as error-prone PCR (epPCR) [Leung, 1989, Cadwell and Joyce, 1992] and DNA shuffling [Stemmer, 1994b] to create libraries of variants with desirable traits. While epPCR can be effectively tuned to control the rate of random mutational frequency [Chen and Arnold, 1993], it can create a needle in a haystack scenario involving huge amounts of screening effort before desirable traits are identified. Furthermore, the evolution strategy must be coupled to a screen capable of achieving the throughput required for success in a realistic time frame. Screening effort is the number of variants required to sample an available sequence space. Focused directed evolution, as discussed in Section 1.2.2, targets residues likely implicated in enzyme catalysis. The aims of this chapter are to use CASTing [Reetz et al., 2005, 2006] for focused directed evolution of *E. coli* threonine aldolase (eTA). We use a known structure of eTA to identify eight residues in close proximity of the ligand binding site which we target for saturation with NDT codons. The developed screen from Chapter 3 is employed to sample libraries of variants for improved stereoselectivity with 2*S*3*R* and 2*S* 3*S* phenylserine (PS) substrates.

4.1.1 The NDT degenerate codon

Degenerate codons can be used to introduce a specific set of amino acids into a protein. While we are able to get all 20 of the canonical amino acids in with a single codon such as NNK or NNS (Table 1.1). This does not give a 20:20 ratio, where 20 codons encode 20 amino acids. What we see are instances where the same amino acid is encoded up to three separate times and others only once. The actual ratio of codons to amino acids with NNK and NNS is 32:20 [Kille et al., 2012].

NDT is the single best degenerate codon for getting a balanced mix of side chain functionality. It has been shown to produce smaller, higher quality libraries than NNK in engineering improved stereoselectivity and activity in enzymes [Reetz et al., 2008]. NDT encodes 12 codons with 12 different amino acids. The available bases, codons, and the corresponding amino acid residues for an NDT codon are shown in Figure 4.1. In this thesis, all the positions listed in Section 4.2.1 were targeted using the NDT condon. Primers used for mutagenesis are shown in Appendix Figure A.2.

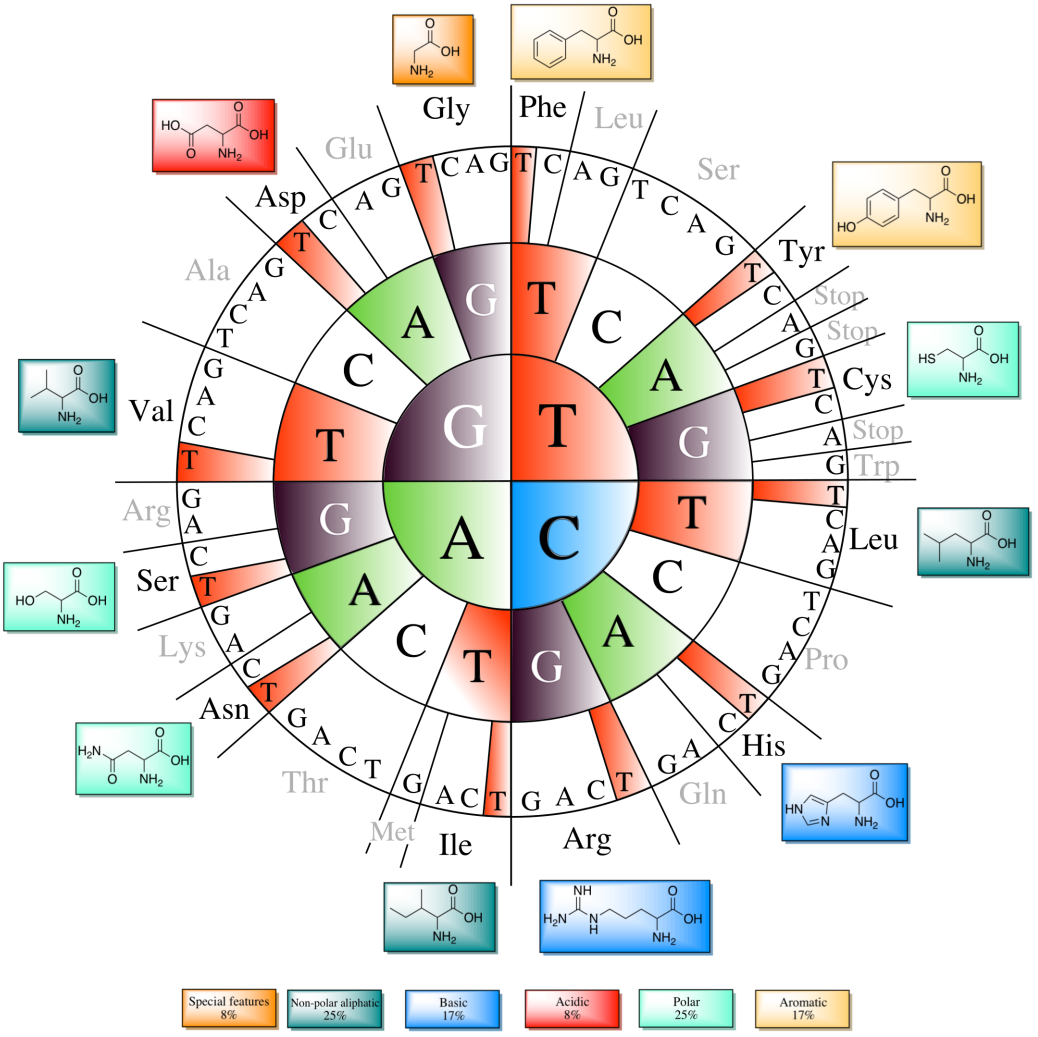


Figure 4.1: Codon wheel highlighted for bases included in the NDT degenerate codon. The four bases encoded by N (A (green), T (red), G (purple) and C (blue)) are in the center of the wheel. Those encoded by D (A T and G) are coloured as before and shown in the second ring. The outer ring highlights thymine (T) in red as this is the only possible base in the codon at this position. Outside the ring, residues capable of insertion are listed by their three letter code in black. The eight not coded for and stop codons are shown in grey. Amino acid structures are also shown and colour coded according to their function.

4.2 Threonine aldolase mutagenesis

Several different L-threonine aldolases (L-TA) have been subject to engineering efforts, giving clues to enzyme selectivity and specificity [di Salvo et al., 2014, Qin et al., 2014, Giger et al., 2012, Gwon et al., 2012]. The structure of an *E. coli* threonine aldolase has been published by di Salvo et al. [di Salvo et al., 2014, Kielkopf and Burley, 2002]. Having this structure provides a useful starting point for the selection of sites to target with focused mutagenesis.

4.2.1 Selection of residues for CASTing of eTA

Eight residues were picked that were in, or close to, the active site of eTA. These were subjected to the CASTing method [Reetz et al., 2006] (Section 1.2.2.1) and paired into four CASTing groups of two amino acids. Residues were chosen based on their location within the protein, and grouped on likely pairwise interactions (Figure 4.2). eTA exists as a homotetramer containing four equivalent active sites [di Salvo et al., 2014](Section 5.2). Each active site contains residues from three of the individual monomers (Figure 4.2).

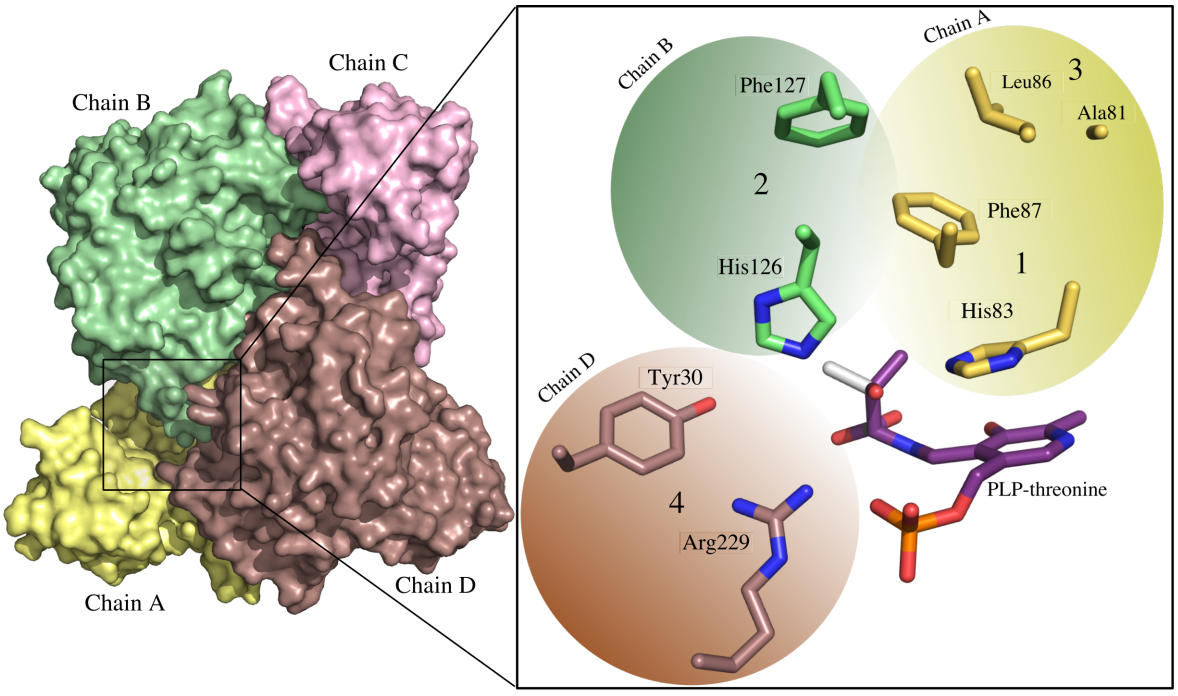


Figure 4.2: eTA shown as the homotetramer composed of chains A (Yellow), chain B (green) chain C (pink) and chain D (brown) from PBD entry 4LNL. The homotetramer was generated from the deposited dimer structure along the 2-fold symmetry axis. Residues from three of the monomer subunits form four equivalent active sites. One of these is expanded to show the selected CASTing residues and their pairwise grouping: Group 1: His83 and Phe87, Group 2: His126 and Phe127, Group 3: Leu86 and Ala81, Group 4: Tyr30 and Arg229. The PLP cofactor in a Schiff base with 2*S*3*S* threonine (white) and 2*S*3*R* threonine (purple) is also shown.

Group 1 consists of His83 and Phe87. Both of these residues have been subject to mutational studies however only for the cleavage of threonine [di Salvo et al., 2014]. His83 is hypothesised as the catalytic base in the retro-aldol cleavage but retains significant activity when mutated to a non-basic residue suggesting this not to be the case [di Salvo et al., 2014]. Phe87 is interestingly the only residue to differ in the active site of eTA and *A. jandaei* TA. Mutating this residue to an alanine in eTA has little effect on the selectivity or

specificity of the enzyme for threonine cleavage [di Salvo et al., 2014]. Group 2 contains His126 and Phe127. His126 is another basic residue in close proximity to the substrate aldol oxygen (Figure 4.2). It could potentiate selectivity, acting as the catalytic base in the retro-aldol mechanism. A H126F variant was found to have increased specificity for 2*S*3*S* threonine cleavage but with no increase in selectivity [di Salvo et al., 2014]. In the *A. jandaei* TA the equivalent histidine (His128) showed improved activity for both L-threonine epimers in a H128Y variant [Qin et al., 2014]. Interestingly, they also found a H128D mutant with a small 1.8-fold increase in selectivity for L-*allo*-threonine with respect to the wild-type enzyme. This variant however, displayed less than 10% of the original activity. Group 3 contains Ala81 and Leu86. This group differs to the others in that the residues are located within the secondary shell of the active site. No published engineering studies have been performed on these residues. Group 4 contains Tyr30 and Arg229. It was recently shown the homologous residue to Arg229 in *A. jandaei* could be changed to an alanine to give improved alanine racemase, and reduced TA, activity [Fesko et al., 2018]. Baik et al. [Gwon et al., 2012] report a Y30C variant as one of four in *Streptomyces coelicolor* TA leading to a five-fold improvement in stereoselectivty for synthesis of L-*threo*-DOPS (2*S*3*R* dihydroxyphenylserine). However, no structure or full gene sequence has been published so it is not possible to confirm whether this is a homologous residue in eTA.

While several of these positions have been subject to mutagenesis studies, none have investigated the affect with PS substrates. Also, the potential synergistic interactions between multiple active site residues remains largely unexplored. One study by Giger et al. [Giger et al., 2012] simultaneously targeted four conserved positions within another *low-specificity* L-TA from *Caulobacter crescentus*. These were targeted by CASTing with NNS degenerate codons. A high-throughput screen was developed whereby a glycine auxotrophic cell strain was created. The knockouts of genes leading to synthesis of glycine in the cell were shown to be lethal to cell growth [Giger et al., 2012]. Subsequent screening of the generated libraries of the TA only yielded colonies still capable of performing the retro-aldol cleavage on either 2*S*3*R* or 2*S*3*S* threonine. The result concluded that a well conserved histidine (homologous to His83 in eTA) was essential for retro-aldol activity as no mutations were tolerated at this position [Giger et al., 2012].

4.2.2 Generation of eTA CASTing libraries

While statistically we can calculate the required screening effort as shown in Section 4.1, this makes a further assumption that each nucleotide base is equally probable of occurring in each CASTing library. For this to be true, there must be no bias in the gene mutagenesis with degenerate primers. Upon inquiry, machine mixed degenerate primers were described as susceptible to "skewing" due to all four (for base N) nucleoside phosphoramidites not hitting the column at the exact same time. Hand mixed degenerate primers, that are not susceptible to this machine bias, were ordered and the occurrence of each base at N, D and T positions following mutagenesis was observed (Figure 4.3).

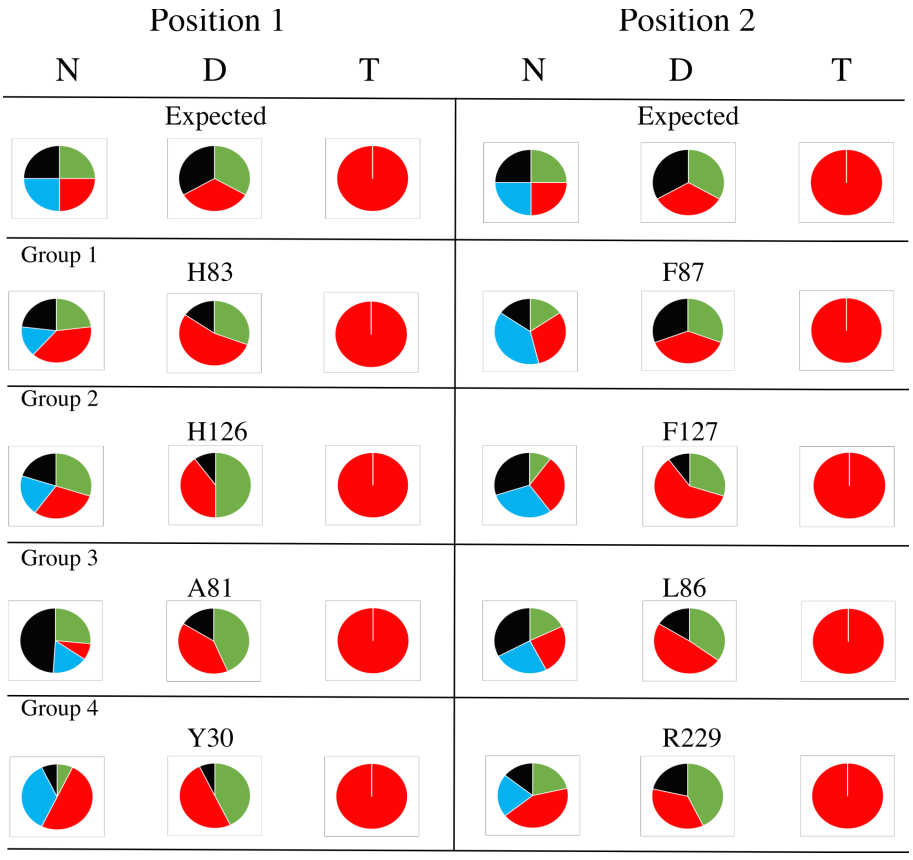


Figure 4.3: NDT sequencing data for CASTing groups 1-4 calculated from the of each base observed for each NDT degenerate codon. 15 colonies for each library were randomly selected and sent for sequencing. Group 4 only contains sequence data for 14 mutants due to one unidentifiable sequence read. The expected ratio for each base is shown in row 1. Nucleotide bases A, T, C and G are shown in green, red, blue and black respectively. At degenerate base N: A, T, C and G are expected to be present in equal amounts (1/4, 1/4, 1/4, 1/4). At degenerate base D: A, T, and G should also be present in equal amounts (1/3, 1/3, 1/3). Base T is used as the final nucleotide in the codon and should be the only nucleotide present. Each amino acid position is highlighted using the single letter alphabet nomenclature and its position in the eTA protein sequence.

The results from sequencing show in all cases, the expected bases corresponding to N, D and T were observed. While next generation sequencing could be used to sequence a larger portion of each library, this was deemed unnecessary given all bases were observed and no unexpected mutations were seen. It was decided, for each group, 540 total colonies would be screened. This would give added confidence that a significant potion of the sequence space for each group was covered.

4.3 CASTing of eTA

Mutagenesis was performed for all libraries as described in Section 2.5.2. Libraries were transformed into competent *E. coli* cells following DpnI digestion and resulting colonies were picked and grown (Section 2.5.3). Reference plates for each library were made and stored at -80^{o}C in case further analysis was required (Section 2.5.4). Crude lysate fractions were prepared (Section 2.5.4) and screened using the developed high-throughput screening method (Section 2.5.5). To give an idea of the developed selection process, an example of the screening and identification procedure is shown (Figure 4.4)[1]. Variants with a threshold activity level of 0.2 ΔAbs_{260}, over a six minute screening period are highlighted in green in the activity plates (Figure 4.4A). To constitute further analysis as a hit, **both** an activity of 0.2 ΔAbs_{260}, and at least a two-fold activity preference for either of the L-PS epimers is required. In well analysis, wells with at least one activity greater than 0.2 ΔAbs_{260} are identified (Figure 4.4B). These identified values are then used to calculate the level of selectivity for each enzyme fraction for the respective L-PS epimer (selectivity analysis) (Figure 4.4C). It is ultimately this selectivity analysis that is used to identify hits for further characterisation. From this point in the thesis all screening activity plates (Figure 4.4A) and the well identifier analysis (Figure 4.4B) will be shown as appendix figures. The calculated selectivity analyses are shown as main text figures. For these, a dark purple colour would be suggestive of a strongly selective variant (up to 10-fold) for 2*S*3*R* PS or dark brown for 2*S*3*S* PS (Figure 4.4C).

[1]The reader might also refer back to Figure 3.15 for clarity of the screening process

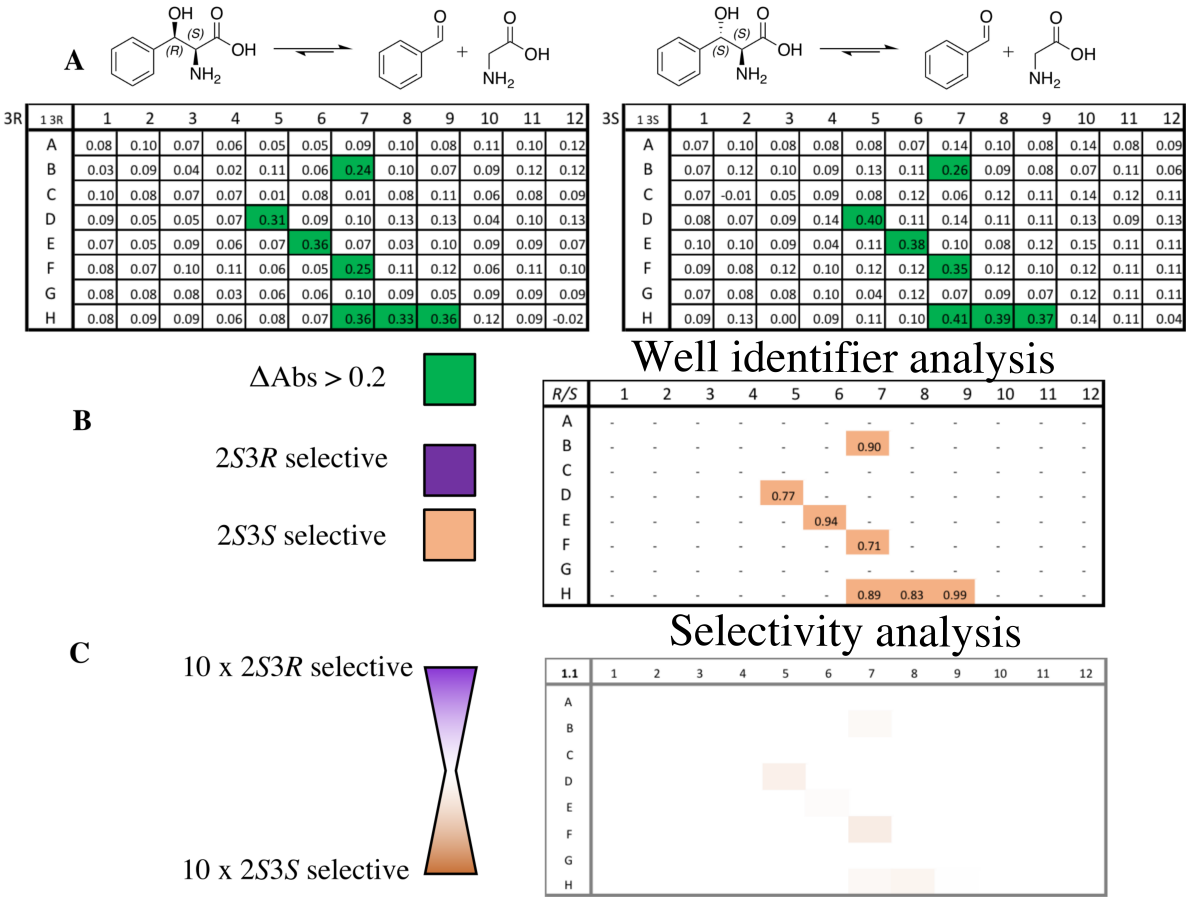


Figure 4.4: From enzyme fraction to identifiable hit: the process of screening one full plate of library variants. Results for the first plate of screened variants for CASTing group 1 ($H83_{NDT}$ $F87_{NDT}$) are shown. A: Two UV-transparent plates are set up in parallel containing 50 mM sodium phosphate pH 8.0 and 10 μL of each prepared lysate fraction. Positive controls containing wild-type eTA are shown in wells H7-9 and negative controls containing empty pKK-mod in wells H10-12. Reactions are started by automated injection of either 2*S*3*R* or 2*S*3*S* PS to each of the respective UV plates to a final concentration of 0.15 mM using a Fluorostar Galaxy plate reader. The change in absorbance at 260 nm is measured over 9 minutes Readings from the final 6 minutes are used to calculate the change in rate to allow the injected wells to first settle. Those wells with a 0.2 ΔAbs_{260} or greater are highlighted in green. The 2*S*3*R* PS screening plate is shown on the left and 2*S*3*S* PS on the right. B: Well identifier analysis: Activity ratios for wells with at least one activity reaching 0.2 ΔAbs_{260} or greater are highlighted (2*S*3*R* / 2*S*3*S*). Those in purple will have greater levels of activity for the 2*S*3*R* PS retro-aldol cleavage and those in brown for 2*S*3*S* PS. C: Selectivity analysis. The level of selectivity observed for each well variant is given by the intensity of purple or brown. Those in a dark purple would be suggestive of at least a 10-fold selective variant for 2*S*3*R* PS. In the example shown no significant active, selective variants are observed.

4.3.1 Group 1, H83$_{NDT}$ F87$_{NDT}$

540 individual colonies from the group 1 library were picked and grown in deep well plates. Crude lysate fractions were prepared and screened. Activity plates and well identifier analysis for group 1 are shown in Appendix Figures A.2-A.4. Calculated selectivity analysis was used to find possible selective variants (Figure 4.5). Several wells met the required criteria for further investigation, all of which suggested variants with increased 2*S*3*S* PS selectivity. It was expected that false positive results would be identified and eliminated in the subsequent follow-up screen. The activity assay described in Section 3.4.1 was performed as a further screen on promising, selective lystate fractions (Figure 4.6A). From this, three variants, 4 F1, 4 G10 and 5 F9, still showed promising 2*S*3*S* PS selectivity. Samples 4 F1, 4 G10 and 5 F9 were purified (Figure 4.6B) and their selectivity again measured with a purified selectivity screen (Figure 4.6C). The selectivity of samples 4 G10 and 5 F9 was subsequently quantified by Michaelis-Menten kinetics (Figure 4.6D) to obtain kinetic parameters k_{cat} and K_m (Table 4.1).

Sequencing showed 4 G10 and 5 F9 to be H83S F87Y and H83C F87Y mutants respectively. Both, showed small increases in the selectivity for 4 G10, the most so with a 2.6-fold selectivity preference (SP) for the 2*S*3*S* PS epimer. Both showed decreases in k_{cat} and K_m with a 30-fold reduction in the 2*S*3*S* PS specificity constants (k_{cat}/K_m) with respect to wild-type eTA.

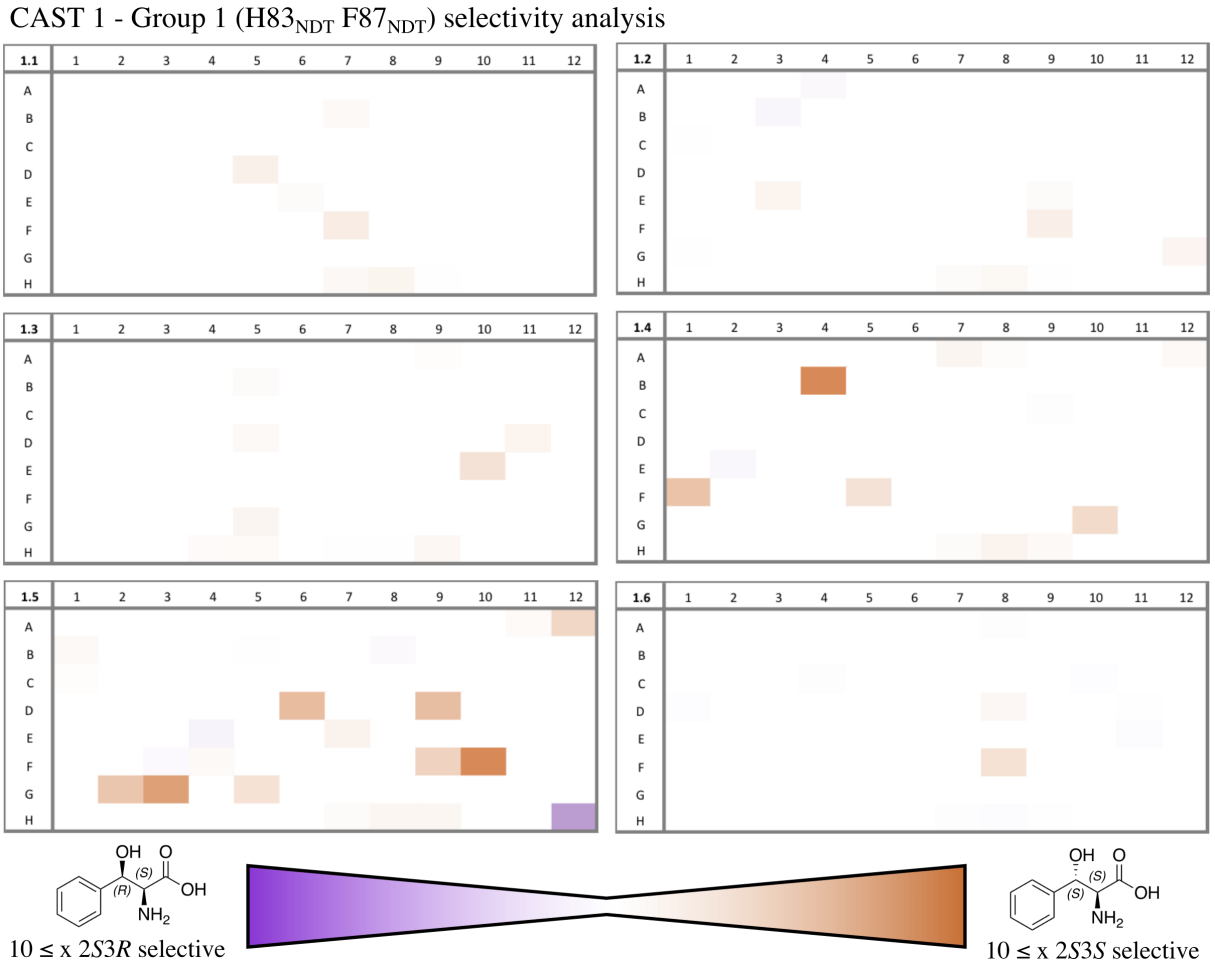


Figure 4.5: Selectivity analysis for CASTing group 1 ($H83_{NDT}$ $F87_{NDT}$). Selectivity of variants reaching the threshold level of activity are highlighted for each plate. A measure of plate selectivity is given by the intensity of the purple or brown colour for 2*S*3*R* and 2*S*3*S* PS respectively. Wells H7-9 contain the wild-type eTA crude lysate positive control and wells H10-12 contain empty plasmid pKK-mod crude lysate negative control. Values in the top left of each plate denote the nomenclature used to identify each screening plate where 1.1 identifies the plate as CASTing group 1 , plate 1 and 1.6 as CASTing group 1, plate 6.

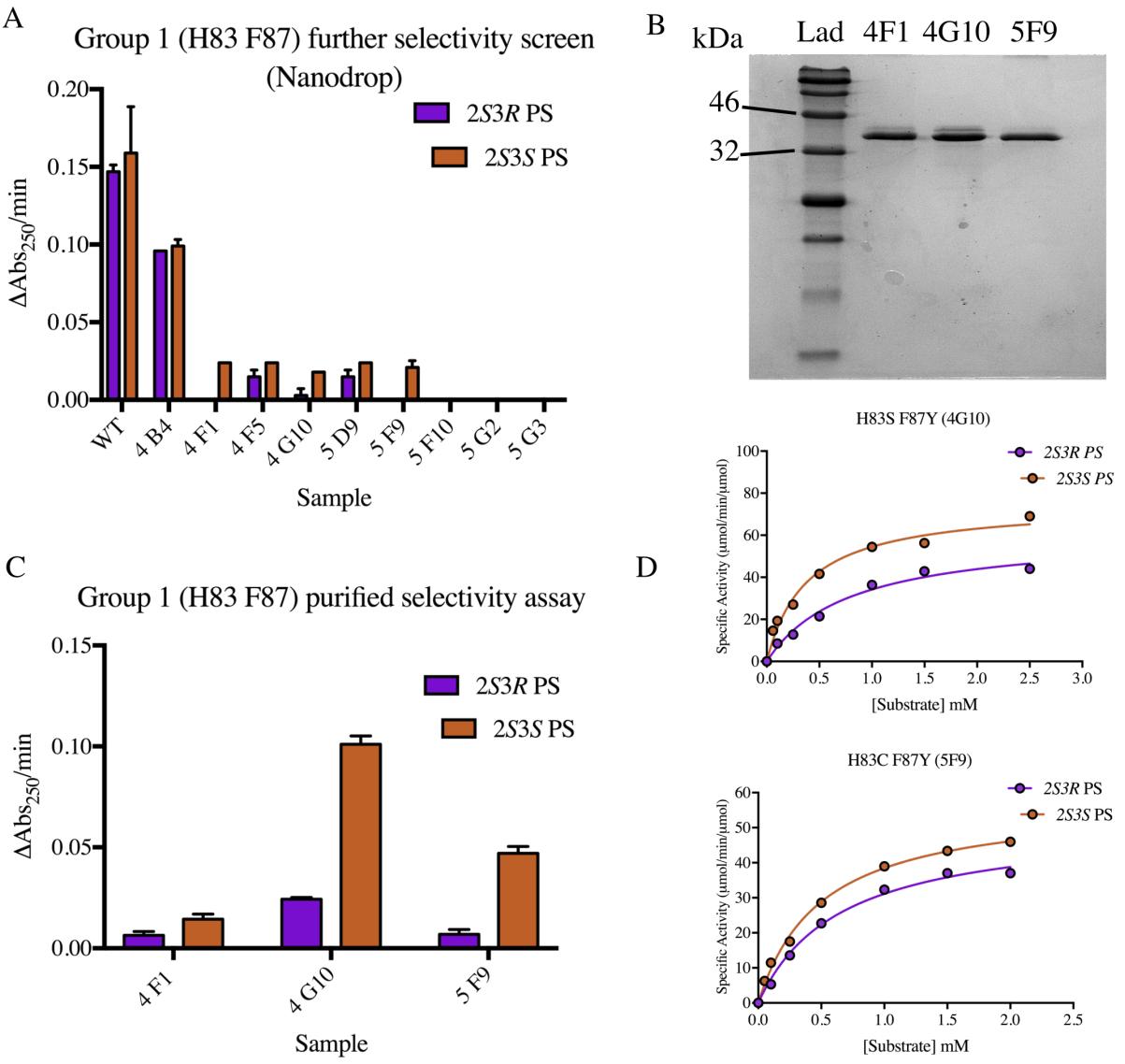


Figure 4.6: A: Nanodrop activity assay of hits showing measurable activity. 70 μL of the lysate fraction was added to 50 mM sodium phosphate buffer pH 8.0 and 50 μM PLP. Reactions were started by addition of either 2*S*3*R* or 2*S*3*S* PS and the rate for each epimer (2*S*3*R* in purple and 2*S*3*S* in brown) retro-aldol cleavage measured over time. B: SDS PAGE gel for the three purified variants, 4 F1, 4 G10 and 5 F9. Each protein was purified using nickel affinity chromatography. Lanes were loaded as shown and run against a protein ladder. All proteins showed strong bands corresponding to masses of overexpressed eTA variants (37 kDa). C: Purified selectivity assay. Reactions contained 50 mM sodium phosphate buffer pH 8.0, 50 μM PLP and 0.1 mM of each of the PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown). Reactions were started by addition of 1 μM of the purified enzymes and the change in absorbance at 250nm measured using a Uvicon spectrophotometer. D: Michaelis-Menten kinetics curves for 4 G10 and 5 F9. Retro-aldol activity for the cleavage of each of the PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown) at a range of concentrations was measured. Each activity measurement was performed at 30°C in 50 mM sodium phosphate buffer pH 8.0 containing 50 μM PLP, 10% (v/v) DMSO and an appropriate amount of enzyme. The change in absorbance at 279 nm was measured and the specific activity calculated using the molar extinction coefficient of 1400 $M^{-1}cm^{-1}$ as described by Liu et al. [Liu et al., 1998].

	2*S3R* Phenylserine			2*S3S* Phenylserine			
Protein	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	SP 3* (*S/R*)
wild-type eTA	92 ± 3	0.016 ± 0.002	5750 ± 740	96 ± 3	0.017 ± 0.002	5650 ± 690	1.0
H83S F87Y (4 G10)	61 ± 5	0.8 ± 0.18	76 ± 18	75 ± 4	0.38 ± 0.07	197 ± 38	2.6 (*S*)
H83C F87Y (5 F9)	51 ± 3	0.65 ± 0.09	78±12	58 ± 2	0.5 ± 0.05	116±12	1.5 (*S*)

Table 4.1: Kinetic parameters for the purified variants from CASTing group 1 ($H83_{NDT}$ $F87_{NDT}$). Kinetics measured for the enzyme catalysed cleavage of 2*S3R* and 2*S3S* PS. The selectivity preference (SP) at the 3 position is calculated from k_{cat}/K_m ratios. For both variants k_{cat} is reduced and K_m increased with respect to wild-type eTA. Sample 4 G10 shows a double H83S F87Y mutant with a 2.6-fold SP for 2*S3S* PS.

4.3.2 Group 2, $H126_{NDT}$ $F127_{NDT}$

The group 2 CASTing library targeted residues H126 and F127. The initial plate screen was performed as with group 1 in Section 4.3.1. Activity plates and well identifier analysis are shown in Appendix Figures A.5-A.7. Selectivity analysis highlighted eight variants that met the criteria for follow up screening (Figure 4.7). Although the Nanodrop assay used for the follow up screen in Section 4.3.1 is suitable for measuring the selectivity of the samples it requires significantly more manual intervention as continuous measurements cannot reliably be made with this model[2]. Increased throughput with the same results can be achieved using a modified version of the plate reader screen and is thus used for all subsequent follow up screening experiments (Section 2.5.6).

The eight lysate samples with promising selectivity were further screened and the respective retro-aldol activity for each of the PS epimers plotted (Figure 4.8). The results showed that the identified "hits" had been false positives as no significant levels of selectivity were observed in the follow up screen. Therefore no further characterisation of these variants was performed.

[2]The Nanodrop2000 model used does not contain the cuvette holder of the 2000c model and measurements are therefore made on the sample pedestal. This requires fresh application of between 1-2.5 μL of the reaction before each measurement is made to avoid sample evaporation.

CAST 1 - Group 2 ($H126_{NDT}$ $F127_{NDT}$) selectivity analysis

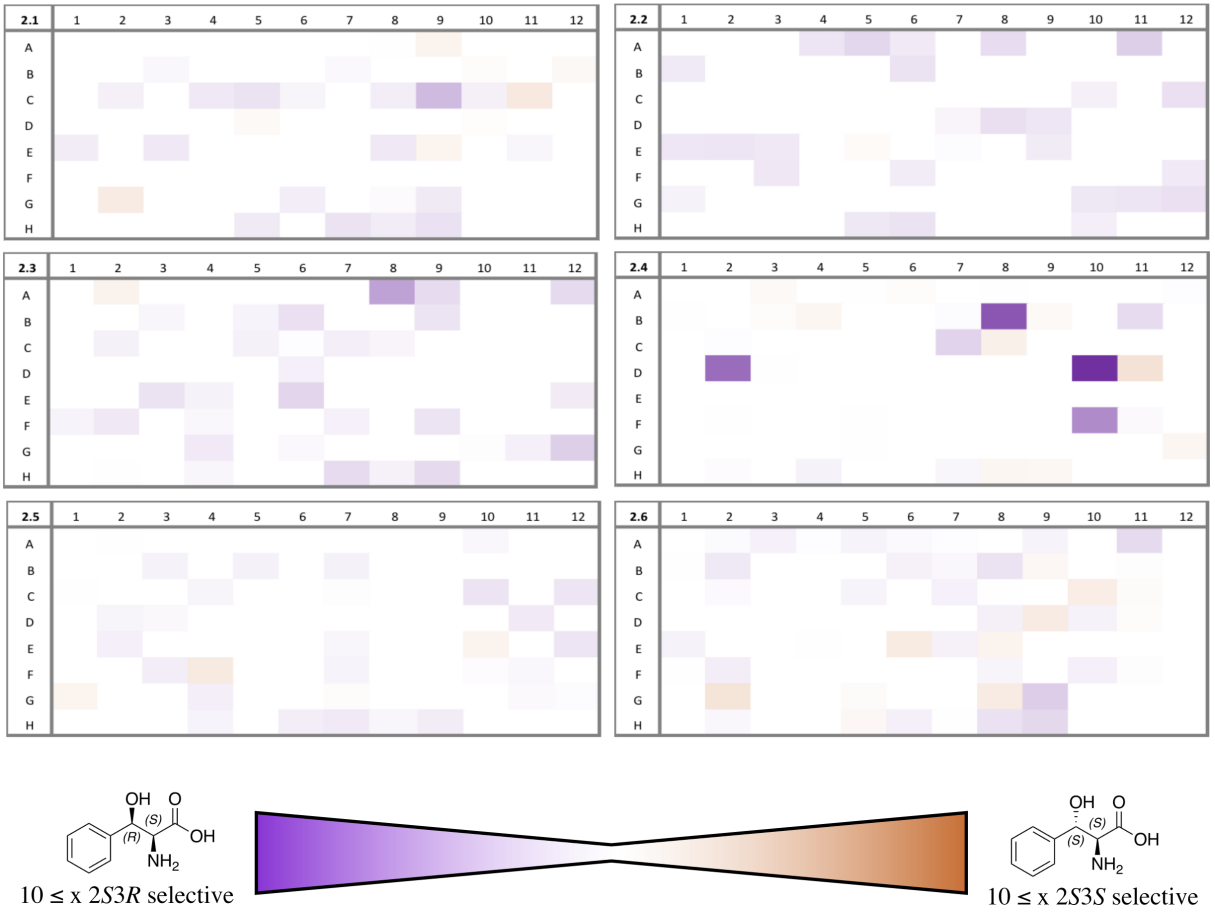


Figure 4.7: Selectivity analysis for CASTing group 2 ($H126_{NDT}$ $F127_{NDT}$). Selectivity of variants reaching the threshold level of activity are highlighted for each plate. A measure of plate selectivity is given by the intensity of the purple or brown colour for 2*S*3*R* and 2*S*3*S* PS respectively.Wells H7-9 contain the wild-type eTA positive control and wells H10-12 contain empty plasmid (pKK-mod) negative control. Values in the top left of each plate denote the nomenclature used to identify each screening plate where 2.1 identifies the plate as CASTing group 2 plate 1 and 2.6 as CASTing group 2 plate 6.

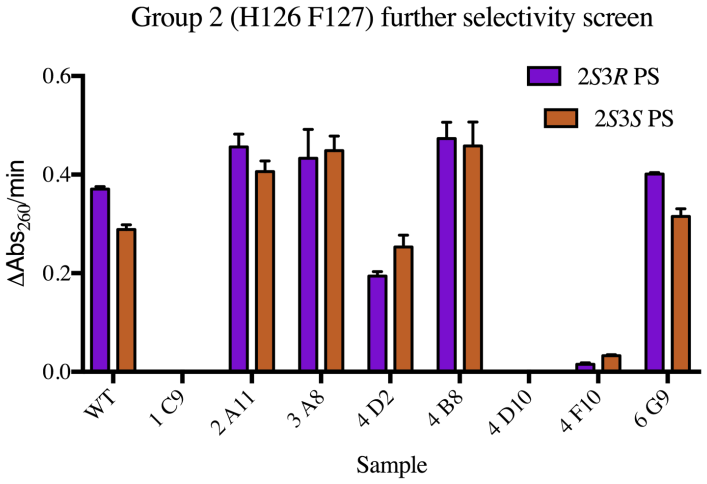


Figure 4.8: Follow up screen for eight group 2 ($H126_{NDT}$ $F127_{NDT}$) variants. Reactions were performed in a Fluorostar Galaxy plate reader. 30 μL of each lysate fraction was added to 50 mM sodium phosphate pH 8.0 containing 50 μM PLP. Reactions were started by addition of the respective PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown) to a screening concentration of 0.15 mM. The total reaction volume for each well was 200 μL. Retro-aldol activity was recorded observing the change in absorbance at 260 nm. None of the measured lysate samples displayed any noteworthy improvements in selectivity.

4.3.3 Group 3, $A81_{NDT}$ $L86_{NDT}$

Group 3 contains residues A81 and L86. These are not active site residues but instead appear behind His 83 and Phe 87 (Figure 4.2). The library was screened for variants with improved selectivity. Activity plates and well identifier analysis are shown in Appendix Figures A.8-A.10. It is clear that the majority of the mutations had little affect on the activity of the enzyme variants, with the vast majority reaching the threshold levels of retro-aldol activity for both the PS epimers. Fifteen variants were identified with levels of activity and selectivity for further screening (Figure 4.9). These were further screened as shown in Figure 4.10. The results suggested no noteworthy improvements in selectivity within the conditions of the screen. Subsequently none of these were taken forward for kinetic characterisation. The results further suggest that the two positions A81 and L86 are unlikely to be key residues in determining stereoselectivity and have little effect on the activity of the enzyme.

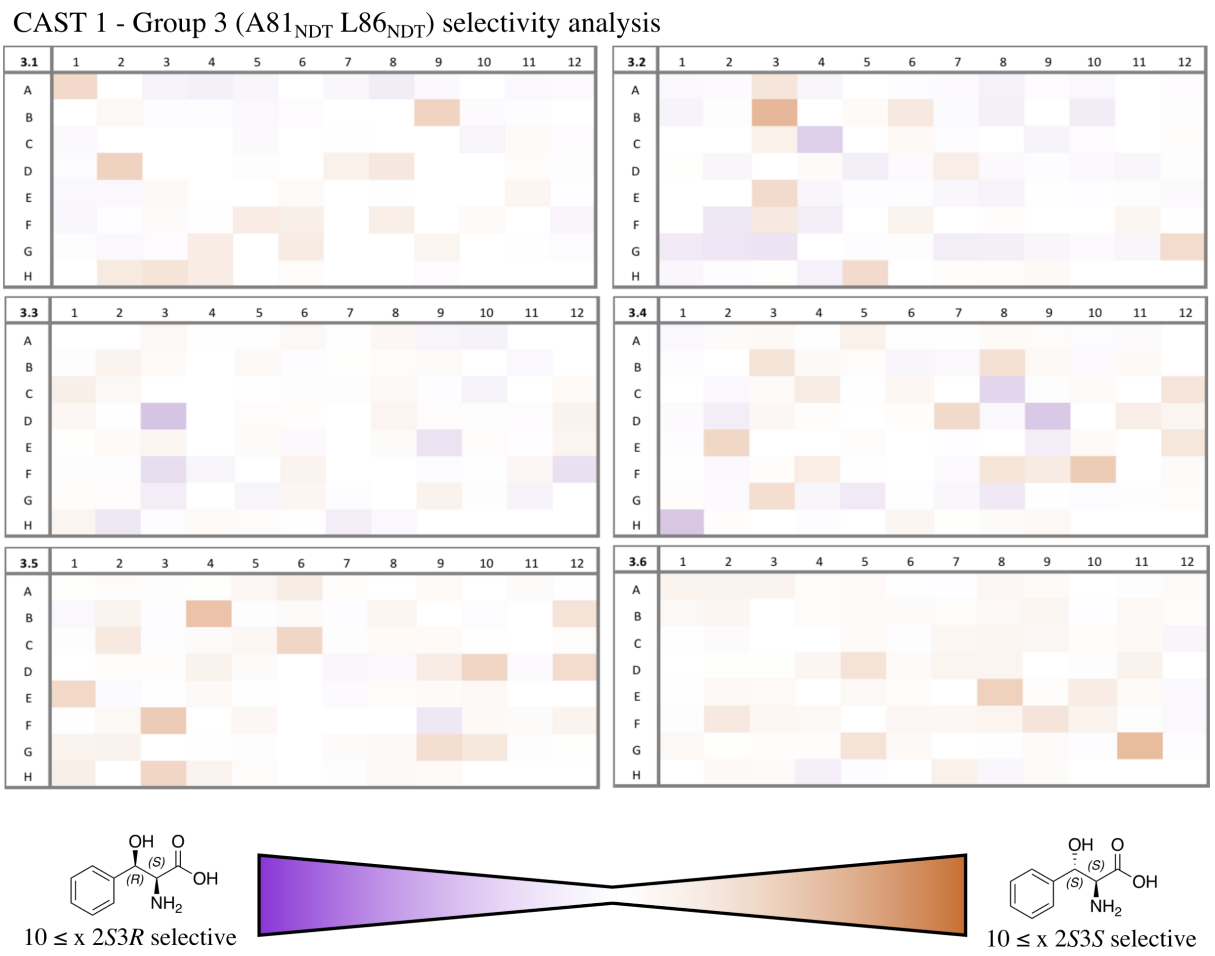


Figure 4.9: Selectivity analysis for CASTing group 3 ($A81_{NDT}$ $L86_{NDT}$). Selectivity of variants reaching the threshold level of activity are highlighted for each plate. A measure of plate selectivity is given by the intensity of the purple or brown colour for 2*S*3*R* and 2*S*3*S* PS respectively. Wells H7-9 contain the wild-type eTA positive control and wells H10-12 contain empty plasmid (pKK-mod) negative control. Values in the top left of each plate denote the nomenclature used to identify each screening plate.

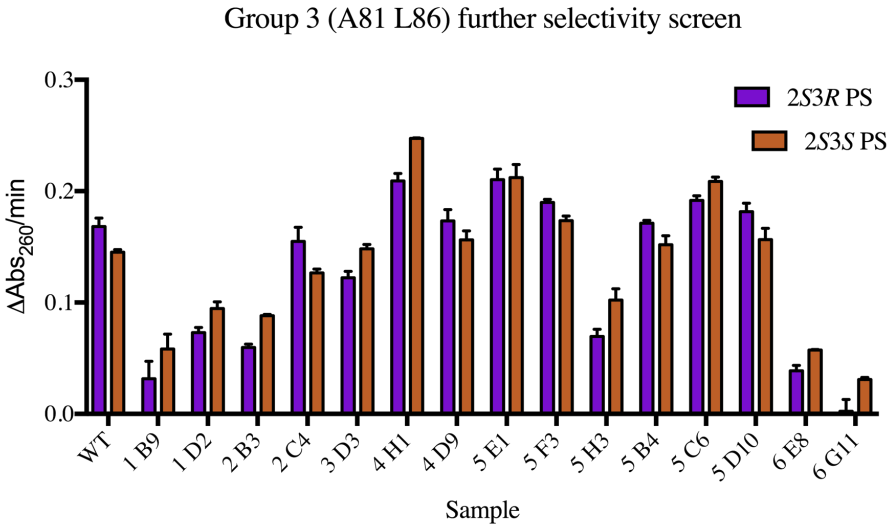


Figure 4.10: Follow up screen for fifteen group 3 ($A81_{NDT}$ $L86_{NDT}$) variants. Reactions were performed in a Fluorostar Galaxy plate reader. 30 μL of each lysate fraction was added to 50 mM sodium phosphate pH 8.0 containing 50 μM PLP. Reactions were started by addition of the respective PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown) to a screening concentration of 0.15 mM. The total reaction volume for each well was 200 μL. Retro-aldol activity was recorded observing the change in absorbance at 260 nm. None of the measured lysate samples displayed any noteworthy improvements in selectivity.

4.3.4 Group 4, $Y30_{NDT}$ $R229_{NDT}$

The final library for screening in the first CASTing cycle was group 4. This group contained residues Y30 and R229. Unlike the other previously screened groups, these residues are not within close sequence proximity despite appearing almost adjacent in the active site (Figure 4.2). The library for this group was created using a multi site-directed mutagenesis kit (Section 2.5.2). Screening of the group 4 library was performed and retro-aldol activity for the enzyme catalysed reactions recorded (Appendix Figures A.11-A.13). 37 hits met the criteria for further analysis (Figure 4.11), significantly more than seen in screening of the previous libraries. These samples were further screened as shown in Figures 4.12A-C. Several of the hits taken forward from plate 4.1 in Figure 4.11 were shown to be false positives; however multiple variants still demonstrated a clear preference for cleavage of 2*S*3*S* PS. This was also seen for the majority of the other variants in the screen. Of the 37 hits initially identified, 24 of those still displayed at least a two-fold preference for the retro-aldol cleavage of the 2*S*3*S* substrate. This suggests that one or both of these targetted positions could be involved in determining stereoselectivity in the reaction. To determine

which of these variants were unique, DNA samples were sent for sequence analysis. This showed 14 of the 24 samples were unique variants (Figure 4.12D). Those initially displaying a slight 2*S*3*R* preference (5 G5 and 6 E10) returned wild-type sequence data. The 14 unique variants were purified (Appendix Figure A.14) and their selectivity measured (Figure 4.13). 12 of the purified variants still displayed at least a two-fold preference for retro-aldol cleavage of the 2*S*3*S* substrate. The occurrence of each residue at position 30 revealed that six residues (plus the wild-type tyrosine) were accepted in the 12 stereoselective variants. In contrast only four (plus the wild-type arginine) were observed at position 229. Interestingly, half of those contained a R229N substitution (Figure 4.13).

CAST 1 - Group 4 ($Y30_{NDT}$ $R229_{NDT}$) selectivity analysis

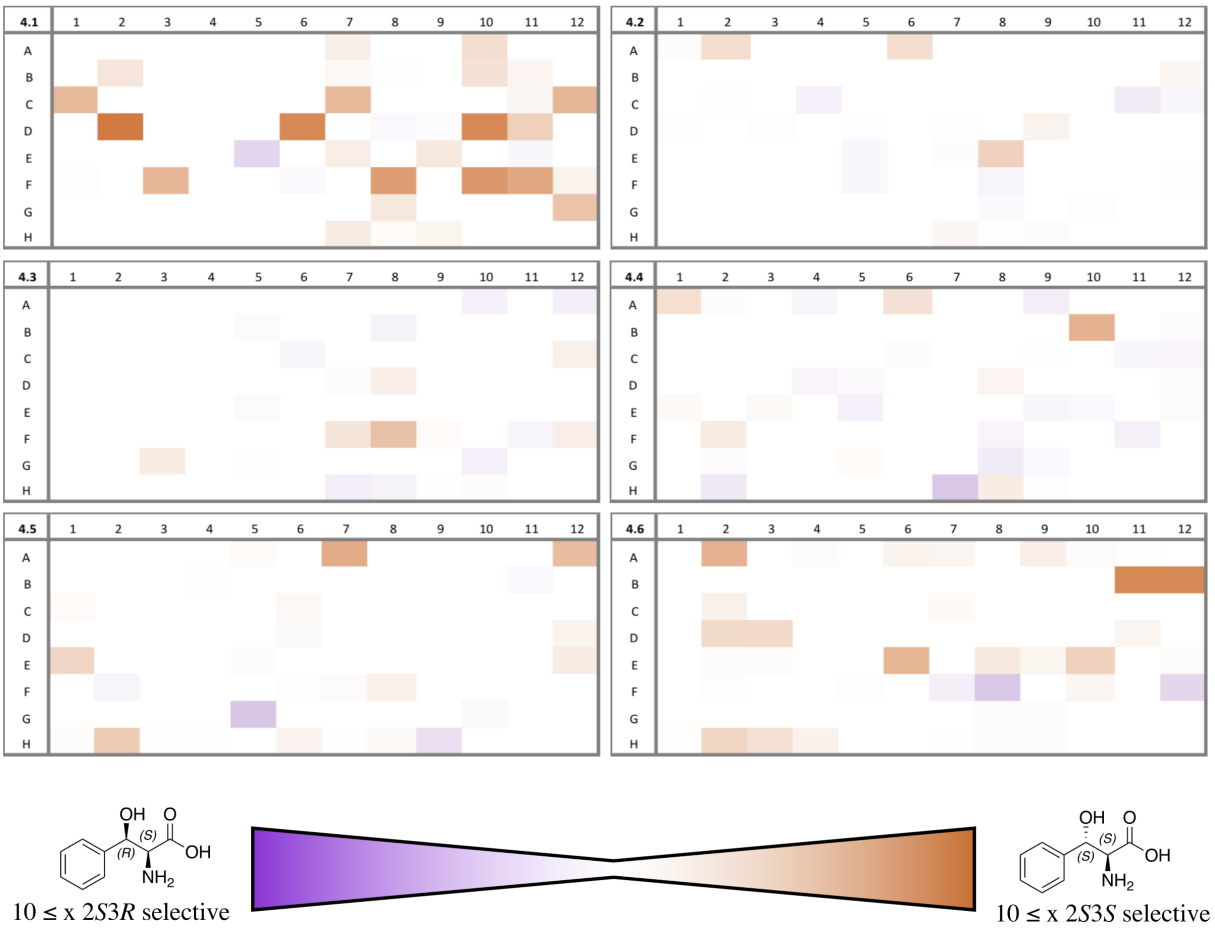


Figure 4.11: Selectivity analysis for CASTing group 4 ($Y30_{NDT}$ $R229_{NDT}$). Selectivity of variants reaching the threshold level of activity are highlighted for each plate. A measure of plate selectivity is given by the intensity of the purple or brown colour for 2*S*3*R* and 2*S*3*S* PS respectively.Wells H7-9 contain the wild-type eTA positive control and wells H10-12 contain empty plasmid (pKK-mod) negative control. Values in the top left of each plate denote the nomenclature used to identify each screening plate. 37 wells met the required criteria for further analysis.

All of these variants had comparably low levels of 2*S*3*R* PS cleavage activity. It was therefore decided that those displaying the highest level of activity for 2*S*3*S* PS would have their selectivity quantified by kinetic characterisation. Kinetic parameters k_{cat} and K_m were determined for the two variants (Figure 4.13C and D) and their values recorded in Table 4.2. 4 B10 (Y30F R229N) had a k_{cat} value 2.4 times greater for the 2*S*3*S* reaction and comparable K_m values, translating to a SP of 3.1 for 2*S*3*S* PS. 3 F8 (Y30R) had similar k_{cat} values for both substrates however a 3.1-fold SP was calculated owing to a 3-fold lower K_m for 2*S*3*S* PS. In comparison to wild-type eTA, both variants had increased k_{cat} values for both of the PS epimers but larger K_m values. As a result the specificity constant's for both variants are between four to five-fold lower than those of wild-type eTA for the retro-aldol

cleavage of 2*S*3*S* PS. Both positions in the group 4 library are thus involved in determining the stereoselectivity of eTA catalysed cleavage of L-PS.

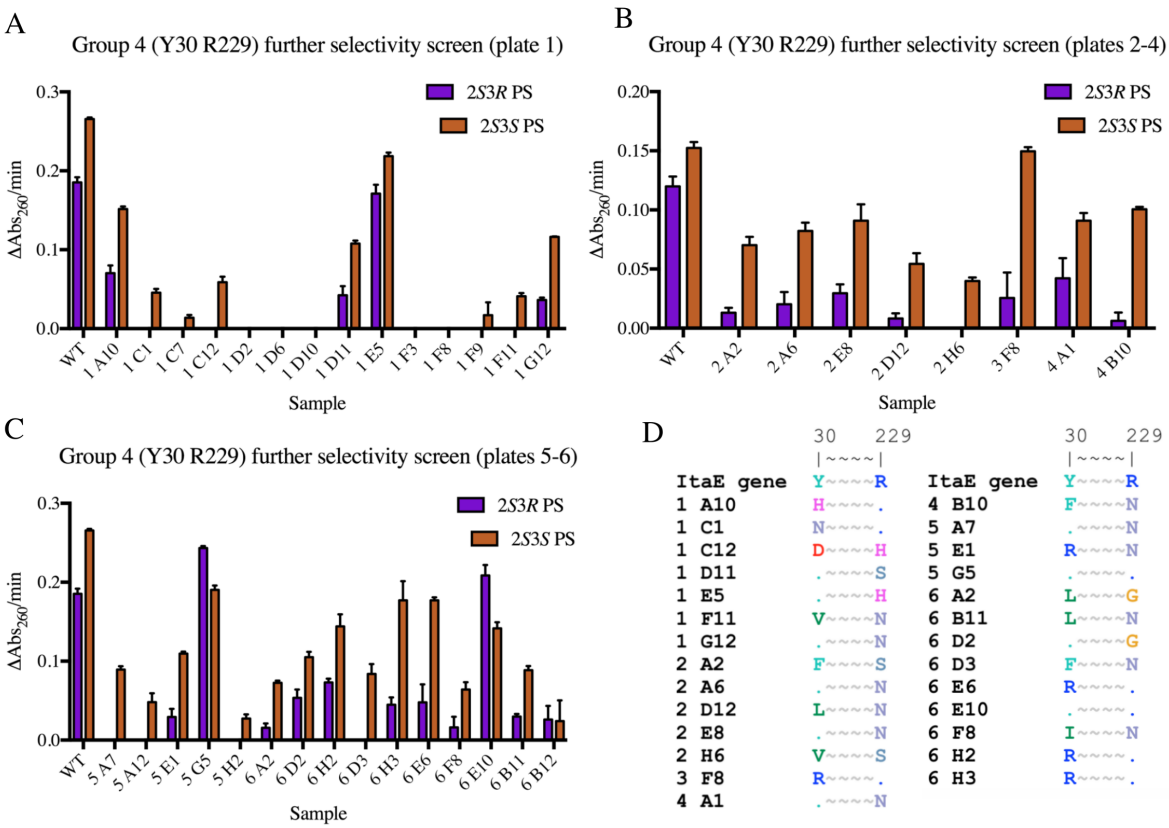


Figure 4.12: A: Follow up screen for variants identified in plate 4.1 of Figure 4.11. 14 of the lysate samples were screened revealing six of these to be notably selective for 2*S*3*S* PS. Reactions were performed in a Fluorostar Galaxy plate reader. 30 μL of each lysate fraction was added to 50 mM sodium phosphate pH 8.0 containing 50 μM PLP. Reactions were started by addition of the respective PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown) to a screening concentration of 0.15 mM. The total reaction volume for each well was 200 μL. Retro-aldol activity was recorded observing the change in absorbance at 260 nm. B: Follow up screen for variants identified in plates 4.2-4.4 of Figure 4.11. All variants showed notable selectivity for 2*S*3*S* PS. Reactions were performed as in A. C: Follow up screen for 15 variants identified in plates 4.5 and 4.6 of Figure 4.11. 11 variants show notable 2*S*3*S* PS selectivity. Reactions were performed as in A. D: Amino acid substitutions at Y30 and and R229 for mutations identified from selective variants. 14 unique variants were observed. No strong preference for a particular residue is observed at position 30 however six of the unique variants contain the R229N mutation (all of which display at least a two-fold preference for 2*S*3*S* PS.

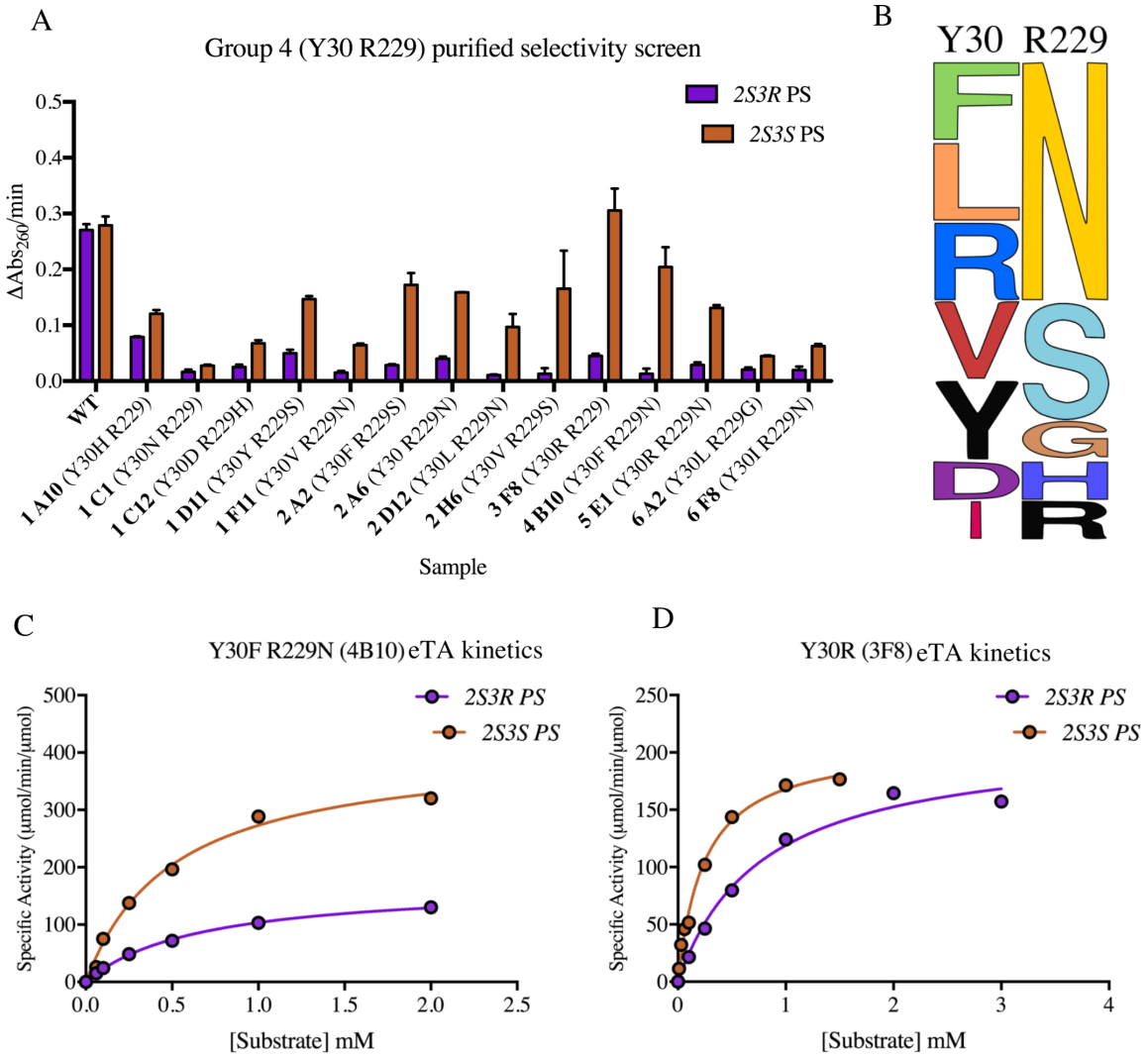


Figure 4.13: A: Purified selectivity assay on unique variants identified from Figure 4.12. Reactions were performed in a Fluorostar Galaxy plate reader. 7 μM of each purified enzyme (50 μg) was added to 50 mM sodium phosphate pH 8.0 containing 50 μM PLP. Reactions were started by addition of the respective PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown) to a screening concentration of 0.15 mM. All enzymes appear to show a two-fold or greater 2*S*3*S* PS stereo-preference except for 1 A10 (Y30H) and 1 C1 (Y30N R229) which show only a slight preference for this reaction. B: The frequency distribution of each amino acid at position 30 and 229 for the 12 2*S*3*S* PS selective variants. No preference for a specific residue appears at position 30 however, at position 229 six of the 12 variants contain the R229N mutation. C and D: Michaelis-Menten kinetics curves for Y30F R229N (4 B10) and Y30R (3 F8) eTA variants. Retro-aldol activity for the cleavage of each of the PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown) at a range of concentrations was measured. Each activity measurement was performed at 30 ^{o}C in 50 mM sodium phosphate buffer pH 8.0 containing 50 μM PLP, 10% (v/v) DMSO and an appropriate amount of enzyme. The change in absorbance at 279 nm was measured and the specific activity calculated using the molar extinction coefficient of 1400 $M^{-1}cm^{-1}$ as described by Liu et al. [Liu et al., 1998].

	2*S*3*R* Phenylserine			2*S*3*S* Phenylserine			
Protein	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	SP 3* (*S*/*R*)
wild-type eTA	92 ± 3	0.016 ± 0.002	5750 ± 740	96 ± 3	0.017 ± 0.002	5650 ± 690	1.0
Y30F R229N (4 B10)	172 ± 4	0.66 ± 0.03	261 ± 13	412 ± 23	0.51 ± 0.07	808 ± 120	3.1 (*S*)
Y30R (3 F8)	212 ± 15	0.79 ± 0.15	268 ± 54	209 ± 8	0.25 ± 0.03	836 ± 105	3.1 (*S*)

Table 4.2: Kinetic parameters for the purified variants from CASTing, group 4 ($Y30_{NDT}$ $R229N_{NDT}$). Kinetics measured for the enzyme catalysed cleavage of 2*S*3*R* and 2*S*3*S* PS. The selectivity preference (SP) at the 3 position is calculated from k_{cat}/K_M ratios for both epimers. Sample 4 B10 contained the double Y30F R229N mutant and displayed a selectivity factor ratio of of 3.1 for 2*S*3*S* PS. Sample 3 F8 also displayed a selectivity factor ratio of 3.1 and contained a single Y30R mutant.

4.4 Rational combinations of CASTing mutations (1)

A greater than three-fold improvement in selectivity had been made following the first round of CASTing for 2*S*3*S* PS cleavage. It was also investigated whether rational combinations of identified variants, in a "best with the best" approach could produce an additive effect on stereoselectivity. This work was carried out in collaboration with an undergraduate student, Sophie Glendinning (Sections 4.4.1 and 4.4.2).

4.4.1 Y30F R229N eTA with additional H83S and F87Y changes

The two best variants found above were H83S F87Y (Group 1, Table 4.1) and Y30F R229N (Group 4, Table 4.2). Primers were ordered to combine these four mutations (Appendix table A.2). Mutagenesis was performed as described in Section 2.5.2 and the recombinant plasmid was transformed into *E. coli* XL10-Gold cells. The correct mutations were confirmed by sequence analysis and the protein was subsequently expressed, purified (Appendix Figure A.21) and sent for accurate mass determination (Appendix Figure A.22). Michaelis-Menten kinetics were measured for the variant as shown in Figure 4.14 and the parameters calculated (Table 4.3).

The quadruple mutant is almost entirely stereoselective for 2*S*3*S* PS with a measured SP of 500. The variant displays a five-fold reduction in k_{cat}, with respect to the wild-type eTA for 2*S*3*S* PS and an 18-fold increase in K_m. This however presents as a 100-fold decrease

in the specificity constant of the enzyme. The quadruple mutant is thus almost entirely stereoselective for the 2*S*3*S* PS retro-aldol cleavage albeit with reduced catalytic efficiency.

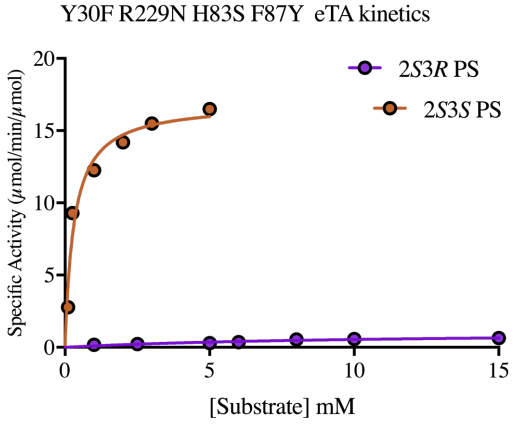


Figure 4.14: Michaelis-Menten kinetics curves for the quadruple Y30F R229N H83S F87Y eTA variant. Retro-aldol activity for the cleavage of each of the PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown) at a range of concentrations was measured. Each activity measurement was performed at 30^{o}C in 50 mM sodium phosphate buffer pH 8.0 containing 50 μM PLP, 10% (v/v) DMSO and an appropriate amount of enzyme. The change in absorbance at 279 nm was measured and the specific activity calculated using the molar extinction coefficient of 1400 $M^{-1}cm^{-1}$ as described by Liu et al. [Liu et al., 1998].

	2*S*3*R* Phenylserine			**2*S*3*S* Phenylserine**			
Protein	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	SP 3* (*S*/*R*)
wild-type eTA	92 ± 3	0.016 ± 0.002	5750 ± 740	96 ± 3	0.017 ± 0.002	5650 ± 690	1.0
Y30F R229N H83S F87Y	1.1 ± 0.3	10.0 ± 4.8	0.11 ± 0.06	17 ± 1	0.31 ± 0.08	55 ± 15	500 (*S*)

Table 4.3: Kinetic parameters for the Y30F R229N H83S F87Y eTA variant. Kinetics measured for the enzyme catalysed cleavage of 2*S*3*R* and 2*S*3*S* phenylserine. The selectivity preference (SP) at the 3 position is calculated from k_{cat}/K_m ratios for both epimers.

4.4.2 Y30F R229N eTA with separate H83S and F87Y changes

To investigate whether both the H83S and the F87Y changes are required for the observed increase in in selectivity, the triple mutants, Y30F R229N H83S and Y30F R229N F87Y were made. Primers were ordered (Appendix Table A.2) and site directed mutagenesis was performed (Section 2.5.2). The recombinant plasmids were transformed into *E. coli* XL10-Gold cells and the proteins were subsequently expressed, purified (Figures A.23 and A.24) and sent for accurate mass determination (Figures A.25 and A.26). Kinetics were performed on both variants (Figure 4.15) and parameters recorded in Table 4.4.

The results suggest a cooperative effect is required between H83S and F87Y to produce the highest level of selectivity when added to the Y30F R229N mutant. No increase in the selectivity is observed in Y30F R229N H83S eTA with respect to the quadruple mutant. The Y30F R229N F87Y variant however, is significantly selective for 2*S*3*S* PS with a SP of 18. The specificity constant of this enzyme for the 2*S*3*S* substrate is eight-fold greater than that of the highly selective Y30F R229N H83S F87Y variant, showing improvements in catalytic efficiency.

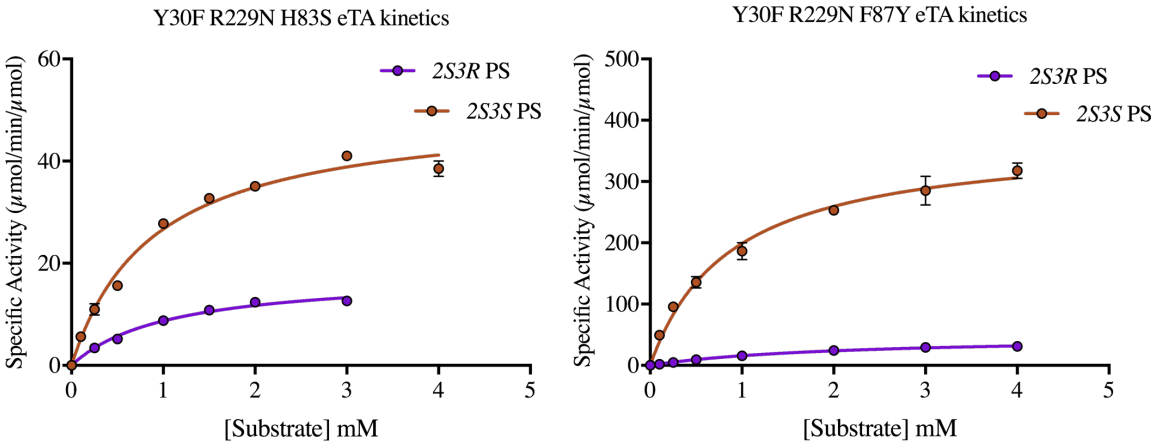


Figure 4.15: Michaelis-Menten kinetics curves for Y30F R229N H83S and Y30F R229N F87Y eTA variants. Retro-aldol activity for the cleavage of each of the PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown) at a range of concentrations was measured. Each activity measurement was performed at 30^{o}C in 50 mM sodium phosphate buffer pH 8.0 containing 50 μM PLP, 10% (v/v) DMSO and an appropriate amount of enzyme. The change in absorbance at 279 nm was measured and the specific activity calculated using the molar extinction coefficient of 1400 $M^{-1}cm^{-1}$ as described by Liu et al. [Liu et al., 1998].

	2*S*3*R* Phenylserine			2*S*3*S* Phenylserine			
Protein	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	SP 3* (*S*/*R*)
wild-type eTA	92 ± 3	0.016 ± 0.002	5750 ± 740	96 ± 3	0.017 ± 0.002	5650 ± 690	1.0
Y30F R229N H83S F87Y	1.1 ± 0.3	10.0 ± 4.8	0.11 ± 0.06	17 ± 1	0.31 ± 0.08	55 ± 15	500 (*S*)
Y30F R229N H83S	18 ± 0.4	1.07 ± 0.12	17 ± 2	50 ± 2	0.88 ± 0.1	57 ± 6	3.4 (*S*)
Y30F R229N F87Y	48 ± 2	2.04 ± 0.15	24 ± 2	373 ± 14	0.88 ± 0.09	424 ± 46	18 (*S*)

Table 4.4: Kinetic parameters for Y30F R229N H83S F87Y eTA and rational combinants. Kinetics measured for the enzyme catalysed cleavage of 2*S*3*R* and 2*S*3*S* PS. The selectivity preference (SP) at the 3 position is calculated from k_{cat}/K_m ratios for both L-PS epimers.

4.5 Iterative CASTing of eTA

Iterative rounds of CASTing (ISM) has proven to be a successful method in developing highly selective enzymes [Reetz et al., 2006, Prasad et al., 2011]. In this approach the best (or second best) variant is taken forward for subsequent mutagenesis with the defined CASTing groups and further screened for the improved characteristic. From Table 4.2, Y30F R229N and Y30R eTA variants both showed greater than a three-fold SP for 2*S*3*S* PS. The double mutant had around a two-fold greater k_{cat} for the 2*S*3*S* PS epimer and further contained the R229N mutation which was present in 46% of the selective, purified variants (Figure 4.13B). It was therefore decided that this mutant (Y30F R229N) would be used as the template for ISM. Figure 4.16 highlights the engineering strategy thus far.

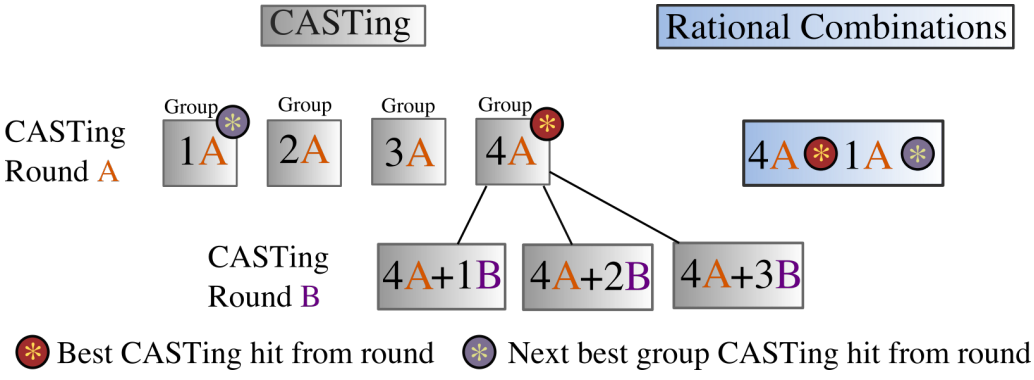


Figure 4.16: The conducted engineering strategy for the directed evolution of eTA. The first round of CASTing (Round A) identified the most selective variants to be from group 4 at positions Y30 and R229. The next best, not from group 4, were identified in group 1 at positions H83 and F87. Rational combinations of these mutations were made, identifying two variants with much improved selectivity for the retro-aldol cleavage of 2*S*3*S* PS. The planned iterative CASTing (CASTing round B) would involve screening of Groups 1-3 with the best mutant identified from group 4.

The same principles relating to screening effort, library generation and hit characterisation were applied as in the first round of CASTing (Section 4.3). One significant change made to the screening process was that "hits" were required to be at least 10-fold selective in selectivity analysis if they are to be further analysed. Rather than using wild-type eTA as the positive control in the screen the selective, Y30F R229N eTA variant would now be used. This would define an internal known level of selectivity.

4.5.1 Group 4 (Y30F R229N) with group 1 ($H83_{NDT}$ $F87_{NDT}$)

The first iterative cycle of eTA CASTing used the Y30F R229N mutant as a template for subsequent screening of the group 1 library (in a group 4 + group 1 manner). The same mutagenic primers as used in the first round of CASTing with group 1, were used (Table A.2) as no overlap with the existing mutations was expected to occur. 540 individual colonies harbouring eTA variants from the 4+1 library were selected (Section 2.5.3) and prepared for screening (Section 2.5.4). The respective 2*S*3*S* and 2*S*3*R* PS retro-aldol activity plates are shown in Figures A.15-A.17. 19 hits were identified from the selectivity analysis as having at least a 10-fold preference for 2*S*3*S* PS (Figure 4.17). These samples were further analysed with a crude lysate selectivity assay (Figure 4.18A). It was however apparent this method could not reliably identify variants with improved selectivity over Y30F R229N eTA due to undetectable levels of 2*S*3*R* PS retro-aldol activity in the lysate samples. The DNA from all 19 mutants was sent for sequence analysis to determine which contained unique mutations. This revealed three unique variants all of which contained an F87 mutant. These were purified (Figure 4.18B) and subject to a purified selectivity assay (Figure 4.18C). All of these showed a strong preference for 2*S*3*S* PS and were kinetically characterised.

All variants showed increased stereoselectivity compared to the starting Y30F R229N eTA. The Y30F R229N F87R (5 A8) variant displayed a SP of 6.7 for 2*S*3*S* PS and the Y30F R229N F87S variant a value of 5.9 for the same reaction. The specificity constants of the enzymes for 2*S*3*S* PS had also further increased and were now only five and seven-fold lower than the wild-type respectively. The three characterised variants all showed similar K_m values between 0.54 and 0.64 mM for 2*S*3*S* PS. This was also seen for K_m values with 2*S*3*R* PS however with increased concentrations (0.85-1.05 mM). The largest difference, owing to

improved selectivity, was seen in the k_{cat} values which were all close to four times greater with 2*S*3*S* PS.

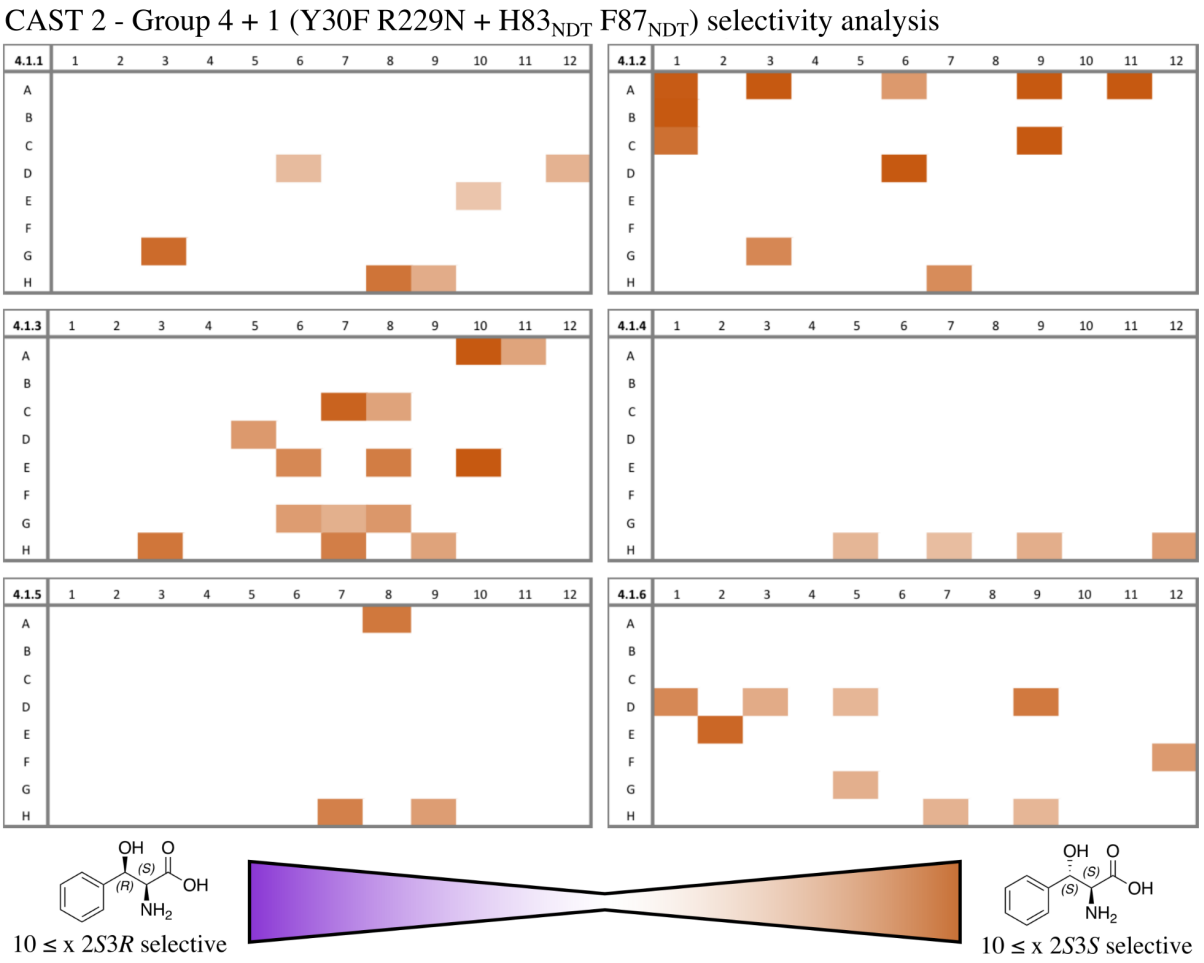


Figure 4.17: Selectivity analysis for CASTing cycle 2: group 4 mutant (Y30F R229N) + group 1 library (H83$_{NDT}$ F87$_{NDT}$). Selectivity of variants reaching the threshold level of activity are highlighted for each plate. A quantitative measure of plate selectivity is given by the intensity of the purple or brown colour for 2*S*3*R* and 2*S*3*S* PS respectively.Wells H7-9 contain the Y30F R229N eTA positive control and wells H10-12 contain empty plasmid (pKK-mod) negative control. Values in the top left of each plate denote the nomenclature used to identify each screening plate where 4.1.1 denotes the plate as CASTing group 4 (Y30F R229N) + CASTing group 1 plate 1. 19 wells met the required criteria for further analysis with at least a 10-fold activity preference for a single diastereomer.

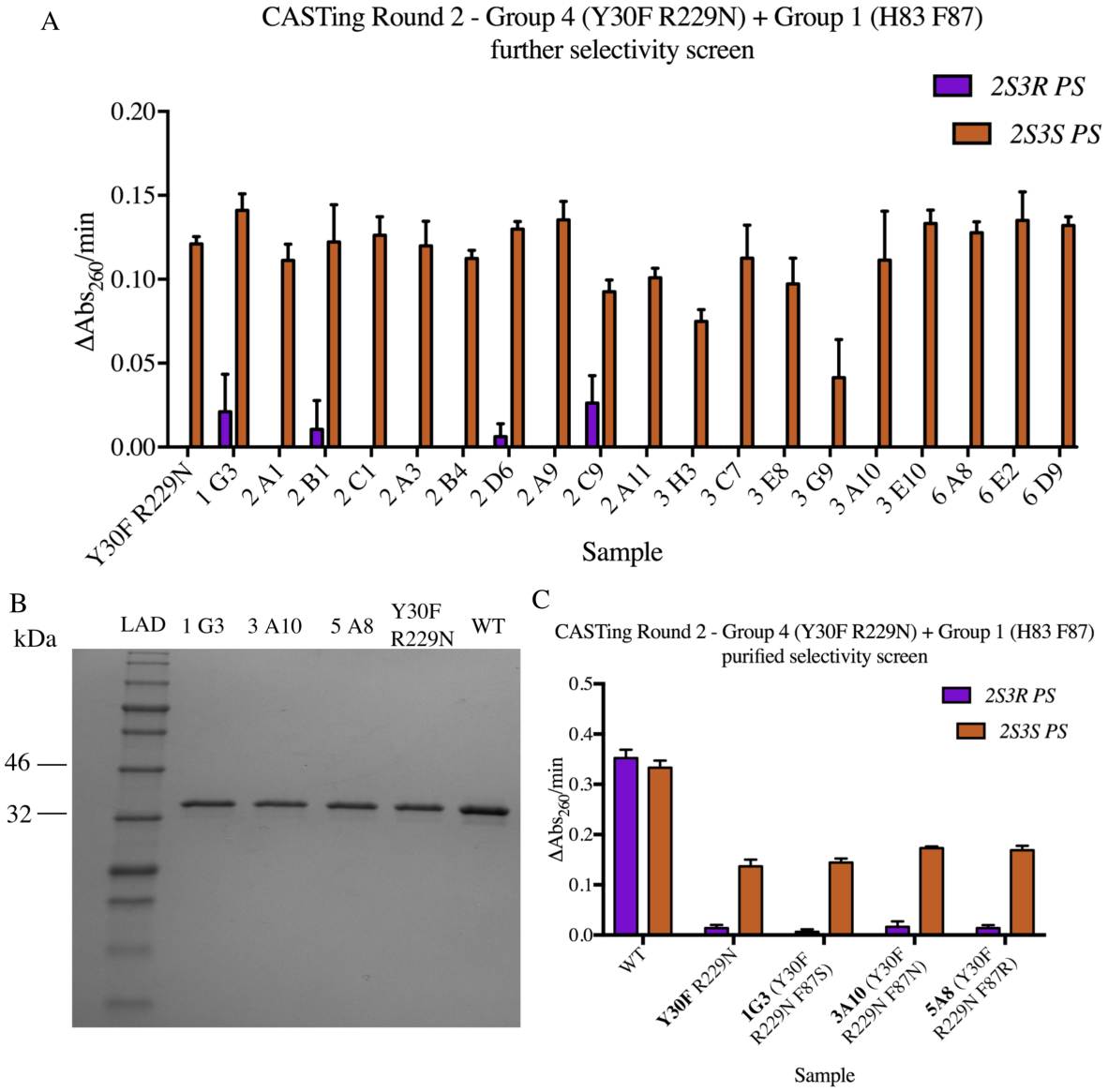


Figure 4.18: A: Follow up selectivity screen for variants identified in Figure 4.17. 19 of the lysate samples were screened, revealing little or no measurable 2*S*3*R* PS retro-aldol activity. Reactions were performed in a Fluorostar Galaxy plate reader. 30 μL of each lysate fraction was added to 50 mM sodium phosphate pH 8.0 containing 50 μM PLP. Reactions were started by addition of the respective PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown) to a screening concentration of 0.15 mM. The total reaction volume for each well was 200 μL. Retro-aldol activity was recorded observing the change in absorbance at 260 nm. B: SDS-PAGE gel for purified variants 1 G3, 3 A10, 5 A8, Y30F R229N and wild-type (WT) eTA. Wells were loaded as labelled. Clear bands were observed for each enzyme around 37 kDa. C: Purified selectivity assay for unique identified variants. All three mutants display similar levels of selectivity. 50 μg of each enzyme was added to 50 mM sodium phosphate buffer, pH 8.0 and 50 μM PLP. Reactions were performed in triplicate and started by injection of the respective PS epimer to a final concentration of 0.15 mM.

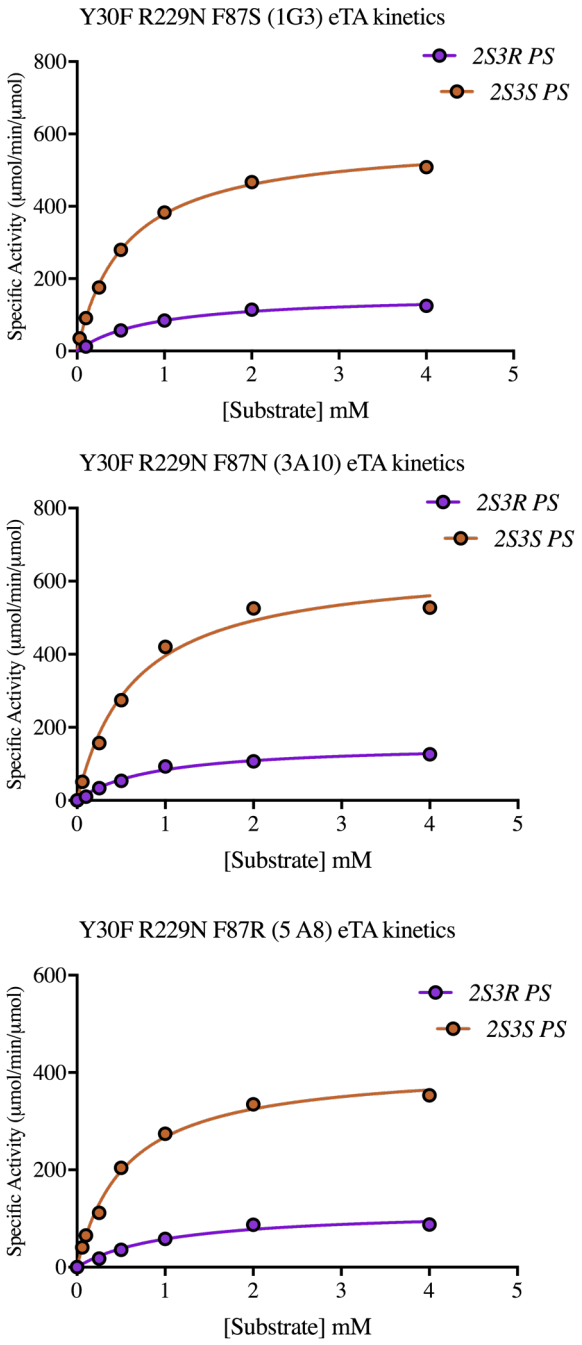


Figure 4.19: Michaelis-Menten kinetics curves for ISM variants from group 4 + group 1: Y30F R229N F87S (1 G3), Y30F R229N F87N (3 A10) and Y30F R229N F87R (5 A8) eTA. Retro-aldol activity for the cleavage of each of the PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown) at a range of concentrations was measured. Each activity measurement was performed at 30^{o}C in 50 mM sodium phosphate buffer pH 8.0 containing 50 μM PLP, 10% (v/v) DMSO and an appropriate amount of enzyme. The change in absorbance at 279 nm was measured and the specific activity calculated using the molar extinction coefficient for benzaldehyde of 1400 $M^{-1}cm^{-1}$ as described by Liu et al. [Liu et al., 1998].

	2*S*3*R* Phenylserine			2*S*3*S* Phenylserine			
Protein	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	SP 3* (*S*/*R*)
wild-type eTA	92 ± 3	0.016 ± 0.002	5750 ± 740	96 ± 3	0.017 ± 0.002	5650 ± 690	1.0
Y30F R229N F87S (1 G3)	156 ± 7	0.85 ± 0.12	184 ± 27	587 ± 8	0.54 ± 0.02	1090 ± 40	5.9 (*S*)
Y30F R229N F87N (3 A10)	155 ± 9	0.85 ± 0.13	182 ± 30	650 ± 38	0.64 ± 0.11	1020 ± 180	5.6 (*S*)
Y30F R229N F87R (5 A8)	119 ± 12	1.05 ± 0.28	113 ± 32	414 ± 13	0.55 ± 0.05	752 ± 72	6.7 (*S*)

Table 4.5: Kinetic parameters for the eTA catalysed cleavage of 2*S*3*R* and 2*S*3*S* phenylserine. The selectivity preference (SP) at the 3 position is calculated from k_{cat}/K_m ratios for both 2*S*3*R* and 2*S*3*S* PS.

4.5.2 Group 4 (Y30F R229N) with Group 2 (H126$_{NDT}$ F127$_{NDT}$)

Iterative CASTing with group 2 involved screening the Y30F R229N H126$_{NDT}$ F127$_{NDT}$ library. NDT primers for group 2 (Appendix Table A.2) were again used to produce the group 4 + group 2 library. Variants were screened for further increases in stereoselectivity. Screening plates are shown in Appendix Figures A.18-A.20 and the calculated selectivity analysis in Figure 4.20. From these, 30 wells displayed the threshold levels of activity and selectivity for further analysis. These were next subject to the selectivity assay screen (Figure 4.21A). As seen in section 4.5.1, levels of 2*S*3*R* PS retro-aldol activity in the lystate fractions was extermely low or unmeasurable. Although not able to reliably identify hits with greater levels of selectivity than the Y30F R229N variant, the follow up screen did confirm that all 30 variants were selective for 2*S*3*S* PS. DNA from these samples was prepared and sent for sequence analysis. Eight unique variants were identified and subsequently purified (Figure 4.21C). Purified samples were again screened for selectivity against the Y30F R229N eTA control (Figure 4.21D). Samples 2 B7, 4 A12, 6 B1 and 6 E12 had their selectivity quantified by Michaelis-Menten kinetics (Figure 4.22). The values for k_{cat}, K_m and the SP are shown in Table 4.6. The two variants, Y30F R229N F127I (4 A12) and Y30F R229N H126R F127L (6 B1) showed similar levels of selectivity for 2*S*3*S* PS to the Y30F R229N mutant. The Y30F R229N F127C (6 E12) eTA shows large increases in K_m for both 2*S*3*R* and 2*S*3*S* PS with similar k_{cat} values resulting in a SP of 2 for the 3*S* epimer. The best hit, and furthermore the best variant identified from CASTing was the Y30F R229N H126R F127H eTA variant. A SP of 9.6 for 2*S*3*S* PS was obtained. Stereoselectivity in this instance is largely due an increase in K_m for binding of the 2*S*3*R* PS epimer. A seven-fold lower K_m for the selective PS substrate was observed as well as a 1.5-fold increase in k_{cat}.

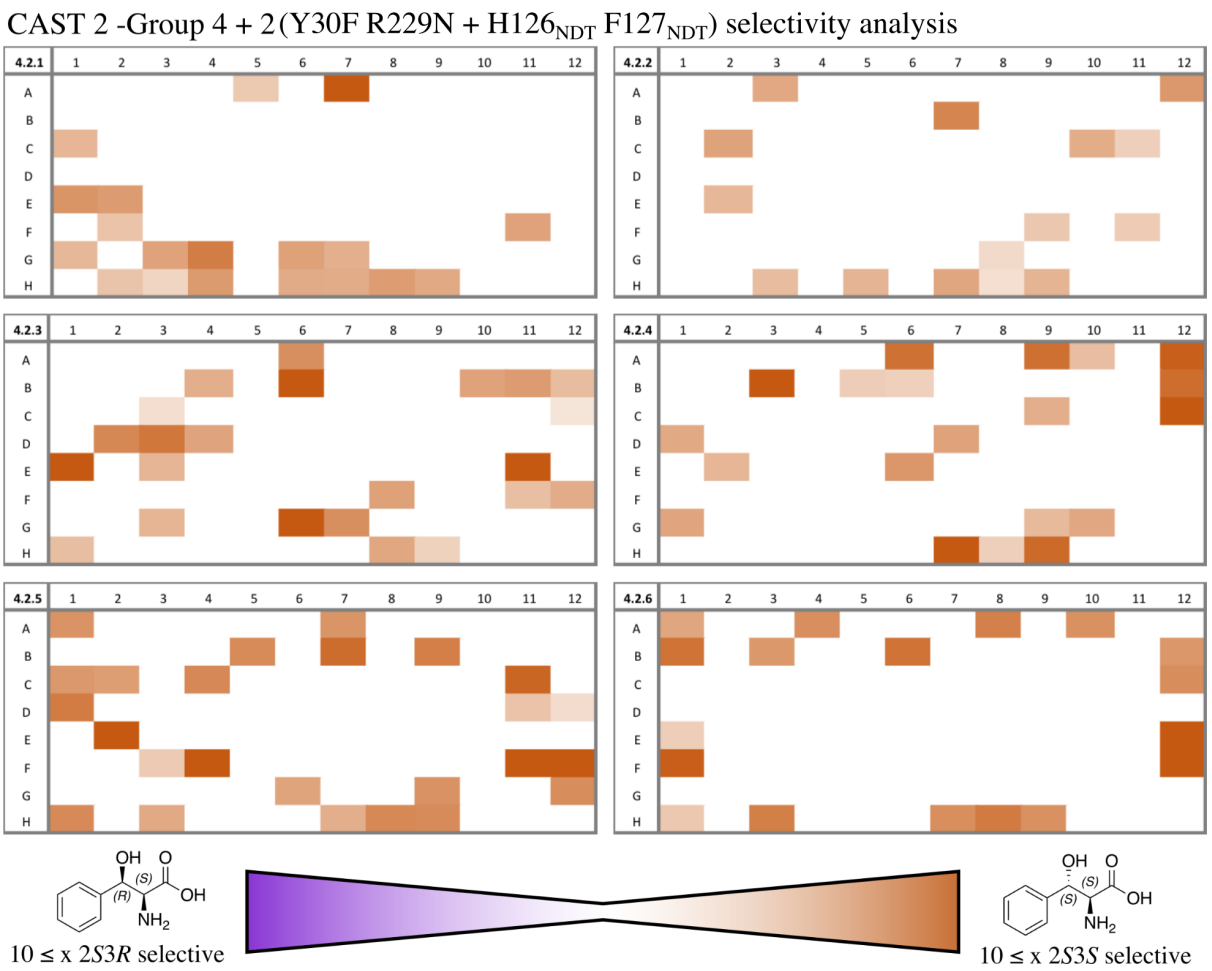


Figure 4.20: Selectivity analysis for CASTing cycle 2: group 4 mutant (Y30F R229N) + group 2 library (H126$_{NDT}$ F127$_{NDT}$). Selectivity of variants reaching the threshold level of activity are highlighted for each plate. A measure of plate stereoselectivity is given by the intensity of the purple or brown colour for 2*S*3*R* and 2*S*3*S* PS respectively.Wells H7-9 contain the Y30F R229N eTA positive control and wells H10-12 contain cells with the empty plasmid (pKK-mod) as a negative control. Values in the top left of each plate denote the nomenclature used to identify each screening plate where 4.2.1 denotes the plate as CASTing group 4 (Y30F R229N) + CASTing group 2 plate 1. 30 wells met the required criteria for further analysis where at least a 10-fold activity preference for a single L-PS epimer was observed.

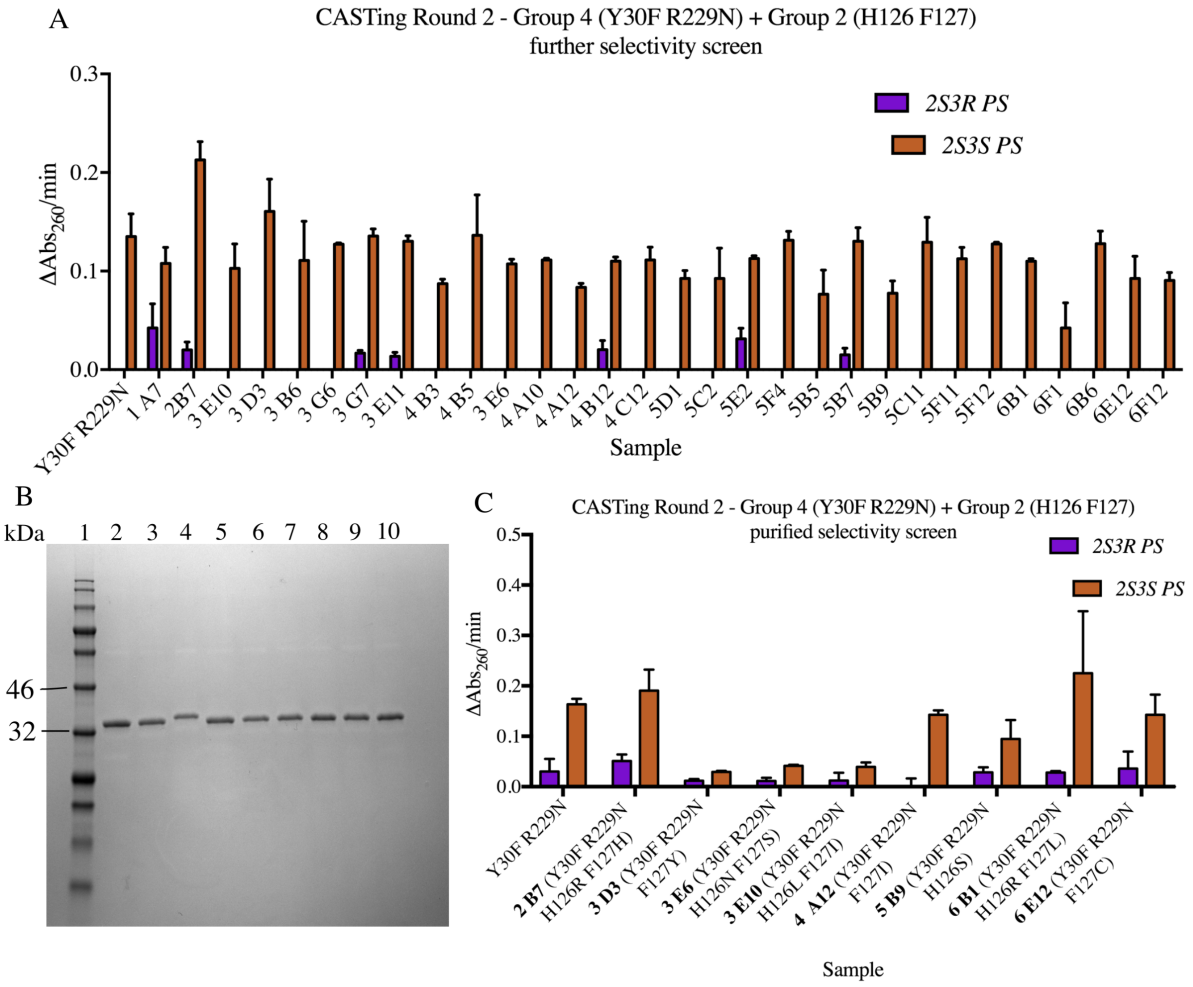


Figure 4.21: A: Follow up screen for variants identified in plate 4.2.1-4.2.6 of Figure 4.20. 15 of the lysate samples met threshold levels of activity and selectivity for further analysis. Reactions were performed in a Fluorostar Galaxy plate reader. 30 μL of each lysate fraction was added to 50 mM sodium phosphate pH 8.0 containing 50 μM PLP. Reactions were started by addition of the respective PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown) to a screening concentration of 0.15 mM. The total reaction volume for each well was 200 μL. Retro-aldol activity was recorded observing the change in absorbance at 260 nm. B: As A for a further 15 samples identified in plates 4.2.5 and 4.2.6 of Figure 4.20. C: SDS PAGE gel of unique variants. Wells loaded as listed: 1: Ladder, 2-10: purified eTA variants: Y30F R229N, 2 B7 (Y30F R229N H126R F127H), 3 D3 (Y30F R229N F127Y), 3 E6 (Y30F R229N H126N F127S), 3 E10 (Y30F R229N H126L F127I), 4 A12 (Y30F R229N F127I), 5 B9 (Y30F R229N H126S), 6 B1 (Y30F R229N H126R F127L), 6 E12 (Y30F R229N F127C) respectively. D: Purified selectivity assay on unique variants identified from A and B. Reactions were performed in a Fluorostar Galaxy plate reader. 7 μM of each purified enzyme (50 μg) was added to 50 mM sodium phosphate pH 8.0 containing 50 μM PLP. Reactions were started by addition of the respective PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown) to a screening concentration of 0.15 mM. Samples 2 B7, 4 A12, 6 B1 and 6 E12 showed the highest levels of 2*S*3*S* PS retro-aldol activity.

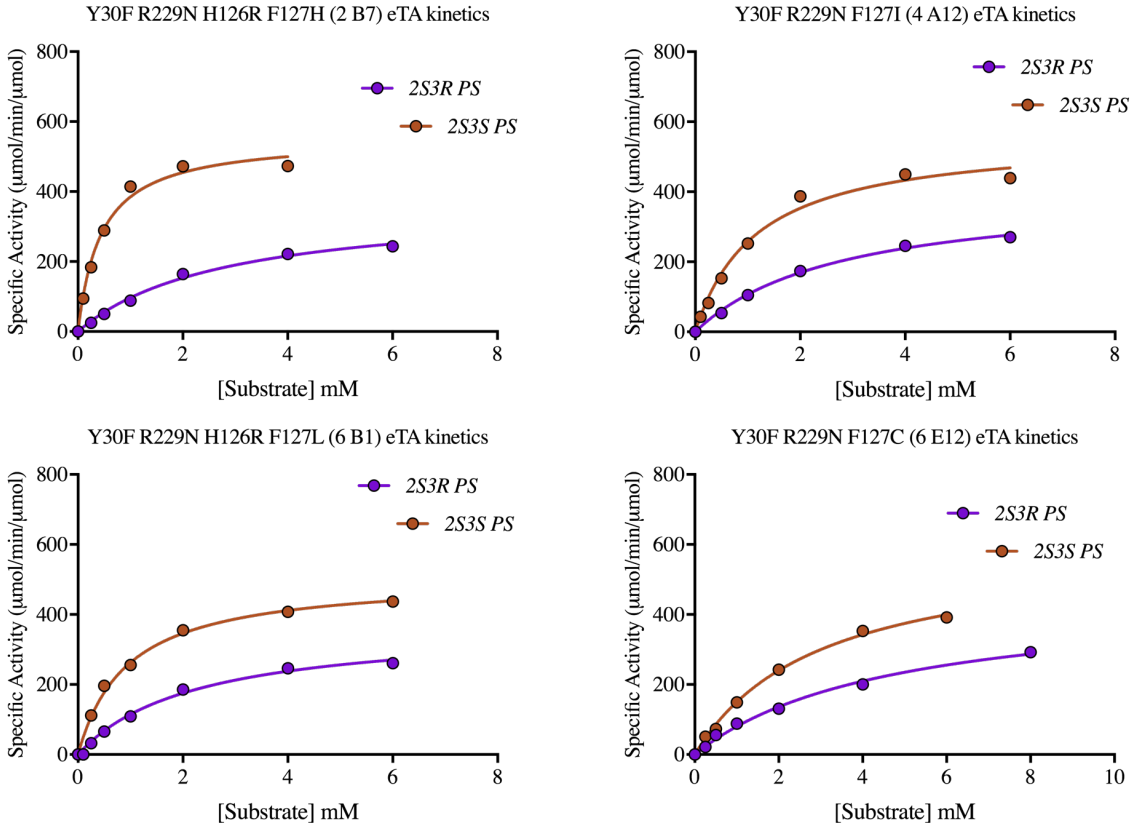


Figure 4.22: Michaelis-Menten kinetics curves for ISM variants from group 4 + group 2: Y30F R229N H126R F127H (2 B7), Y30F R229N F127I (4 A12), Y30F R229N H126R F127L (6 B1) and Y30F R229N F127C (6 E12) eTA. Retro-aldol activity for the cleavage of each of the PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown) at a range of concentrations was measured. Each activity measurement was performed at 30°C in 50 mM sodium phosphate buffer pH 8.0 containing 50 μM PLP, 10% (v/v) DMSO and an appropriate amount of enzyme. The change in absorbance at 279 nm was measured and the specific activity calculated using the molar extinction coefficient for benzaldehyde of 1400 $M^{-1}cm^{-1}$ as described by Liu et al. [Liu et al., 1998].

	2*S*3*R* Phenylserine			**2*S*3*S* Phenylserine**			
Protein	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	SP 3* (*S*/*R*)
wild-type eTA	92 ± 3	0.016 ± 0.002	5750 ± 740	96 ± 3	0.017 ± 0.002	5650 ± 690	1.0
Y30F R229N H126R F127H (2 B7)	368 ± 24	2.81 ± 0.41	131 ± 21	555 ± 25	0.44 ± 0.07	1260 ± 210	9.6 (*S*)
Y30F R229N F127I (4 A12)	405 ± 20	2.80 ± 0.31	144 ± 18	560 ± 31	1.17 ± 0.19	478 ± 82	3.3 (*S*)
Y30F R229N H126R F127L (6 B1)	371 ± 24	2.24 ± 0.34	166 ± 27	507 ± 28	0.93 ± 0.16	545 ± 99	3.3 (*S*)
Y30F R229N F127C (6 E12)	455 ± 33	4.69 ± 0.67	97 ± 16	598 ± 28	2.99 ± 0.30	200 ± 22	2.1 (*S*)

Table 4.6: Kinetic parameters for the eTA catalysed cleavage of 2*S*3*R* and 2*S*3*S* phenylserine . The selectivity preference (SP) at the 3 position is calculated from k_{cat}/K_m ratios for both 2*S*3*R* and 2*S*3*S* PS.

4.5.3 Further investigation of the Y30F R229N H126R F127H eTA variant

The Y30F R229N H126R F127H eTA variant identified during the iterative cycle of CASTing is close to 10-fold selective for 2*S*3*S* PS. Although both the triple mutants, Y30F R229N H126R and Y30F R229N F127H should have been present in the CASTing library, it is possible that these variants may not have been screened. We therefore investigated whether the increase in stereoselectivity owes to the H126R or F127H change, or whether both are required to produce a synergistic effect.

4.5.3.1 Y30F R229N H126R and Y30F R229N F127H

Primers were ordered to make the Y30F R229N H126R and Y30F R229N F127H eTA variants (Appendix Table A.2). Mutagenesis was performed (Section 2.5.2) and recombinant plasmids were transformed into *E. coli* XL10-Gold (Section 2.3.5). DNA was prepared and sent for sequencing to confirm the correct mutations had been made. Following successful results from sequencing the proteins were expressed and purified (Appendix Figure A.27) and sent for accurate mass determination (Appendix Figures A.28 and A.29). Kinetics were performed on both variants (Figure 4.23) and the parameters k_{cat} and K_m recorded in Table 4.7. The results showed that the 10-fold level of 2*S*3*S* PS selectivity can only be achieved when the quadruple mutant, combining both H126R and F127H mutations, is present. Despite reductions in the stereoselectivity of the Y30F R229N H126R and Y30F R229N F127H variants, significant improvements in the k_{cat} values for both 2*S*3*R* and 2*S*3*S* PS are observed to levels over 25 times that of wild-type eTA. The Y30F R229N F127H variant displays a specificity constant for 2*S*3*S* PS of 2400 $min^{-1}mM^{-1}$, 2.4 times lower than that of wild-type eTA and three times greater than that of the Y30F R229N variant. This suggests a significant improvement in the catalytic ability of the enzyme, with respect to the best first round CASTing mutant, while retaining similar levels of stereoselectivity.

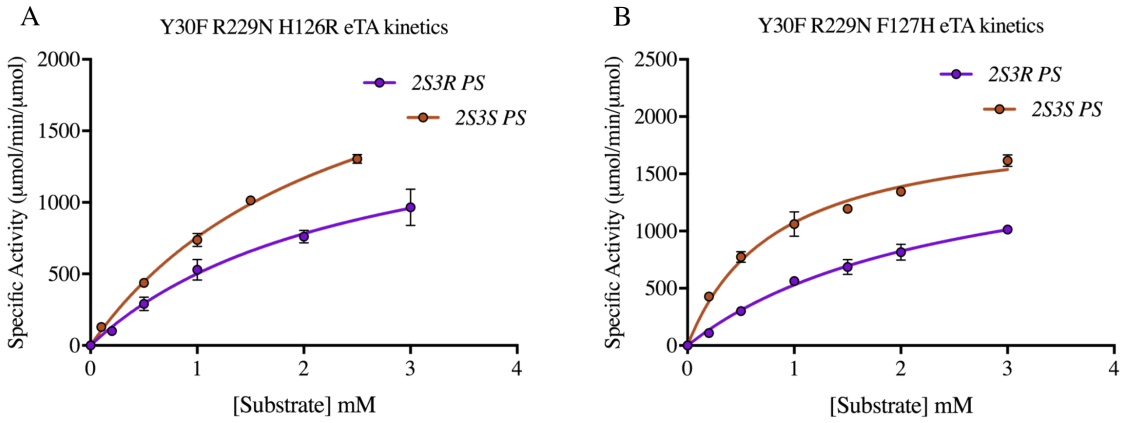


Figure 4.23: Michaelis-Menten kinetics curves for Y30F R229N H126R and Y30F R229N F126H eTA. Retro-aldol activity for the cleavage of each of the PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown) at a range of concentrations was measured. Each activity measurement was performed at 30^{o}C in 50 mM sodium phosphate buffer pH 8.0 containing 50 μM PLP, 10% (v/v) DMSO and an appropriate amount of enzyme. The change in absorbance at 279 nm was measured and the specific activity calculated using the molar extinction coefficient for benzaldehyde of 1400 $M^{-1}cm^{-1}$ as described by Liu et al. [Liu et al., 1998].

	2*S*3*R* Phenylserine			**2*S*3*S* Phenylserine**			
Protein	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	SP 3* (*S*/*R*)
wild-type eTA	92 ± 3	0.016 ± 0.002	5750 ± 740	96 ± 3	0.017 ± 0.002	5650 ± 690	1.0
Y30F R229N H126R	1770 ± 250	2.53 ± 0.63	699 ± 199	2560 ± 170	2.38 ± 0.26	1080 ± 140	1.5 (*S*)
Y30F R229N F127H	1860 ± 150	2.51 ± 0.36	740 ± 122	1960 ± 90	0.82 ± 0.10	2390 ± 310	3.2 (*S*)

Table 4.7: Kinetic parameters for the eTA catalysed cleavage of 2*S*3*R* and 2*S*3*S* phenylserine. The selectivity preference (SP) at the 3 position is calculated from k_{cat}/K_M ratios for both 2*S*3*R* and 2*S*3*S*

4.5.4 Rational combinations of CASTing mutations (2)

Through CASTing we have achieved and almost 10-fold stereoselective eTA for 2*S*3*S* PS. It was decided that no further screening of CASTing libraries would be performed. The developed high-throughput screen had shown to be successful in identifying selective variants, however once a level of selectivity had been achieved it became difficult to discriminate between further improved hits without full kinetic characterisation. Therefore, the engineering process for this project would be finished by making a final rational combination of the two best, separate group hits from the iterative CASTing round (Figure 4.24).

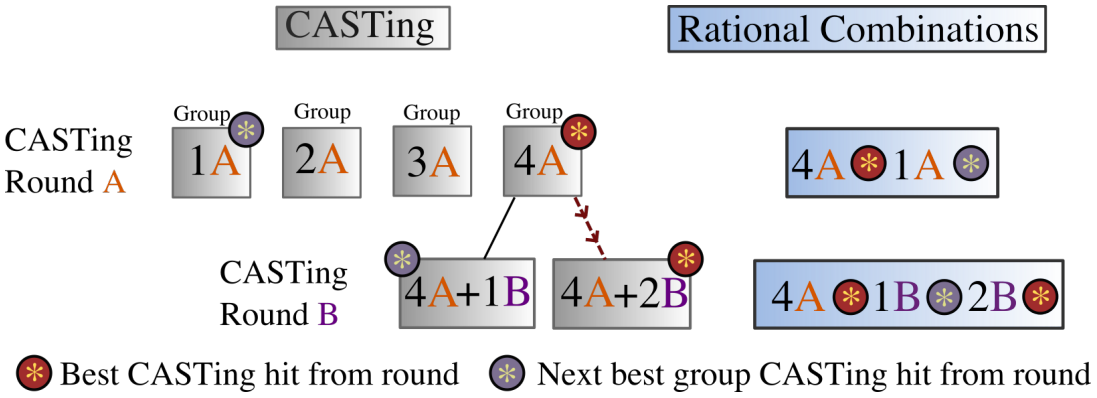


Figure 4.24: The conducted engineering strategy for the directed evolution of eTA. The first round of CASTing identified the most selective variants to be from group 4 at positions Y30 and R229. The next best, not from group 4, were identified in group 1 at positions H83 and F87. Rational combinations of these mutants were made identifying two variants with much improved selectivity for the retro-aldol cleavage of 2*S*3*S* PS. ISM (CASTing round B) identified further improvements in stereoselctivty. The most selective variant was identified from group 2 which contained a Y30F R229N H126R F127H variant. The next best variant was identified in group 1 from Y30F R229N F87R eTA. The rational combination for these mutations would create a quintuple Y30F R229N F87R H126R F127H variant.

In Section 4.4.1, we showed rational combinations of the best variants from group 1 and 4 resulted in a close to 20-fold stereoselective variant for 2*S*3*S* PS. Therefore, by the same process, the 6.7-fold 2*S*3*S* selective Y30F R229N F87R mutation would be combined with the 10-fold 2*S*3*S* selective Y30F R229N H126R F127H mutation as shown in Figure 4.24. Primers to create the quintuple Y30F R229N F87R H126R F127H mutant were ordered (Appendix Table A.2) and mutagenesis performed. The recombinant plasmid was transformed into *E. coli* XL10-Gold cells and single colonies were sent for sequencing. Following successful cloning of the mutant plasmid, the protein was expressed, purified (Figure A.30) and sent for accurate mass analysis (Figure A.31). Michaelis-Menten kinetics were performed as shown in Figure 4.25 and the parameters k_{cat} and K_m recorded in Table 4.8. The results showed no increase in the stereoselectivity of the enzyme and a poor SP of 1.8 (*3S*).

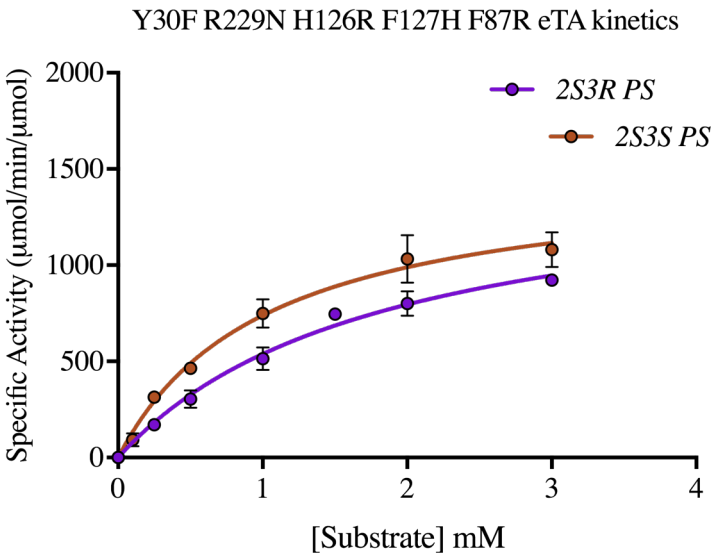


Figure 4.25: Michaelis-Menten kinetics curves for Y30F R229N H126R F127H F87R eTA. Retro-aldol activity for the cleavage of each of the PS epimers (2*S*3*R* in purple and 2*S*3*S* in brown) at a range of concentrations was measured. Each activity measurement was performed at 30^{o}C in 50 mM sodium phosphate buffer pH 8.0 containing 50 μM PLP, 10% (v/v) DMSO and an appropriate amount of enzyme. The change in absorbance at 279 nm was measured and the specific activity calculated using the molar extinction coefficient for benzaldehyde of 1400 $M^{-1}cm^{-1}$ as described by Liu et al. [Liu et al., 1998].

	2*S*3*R* Phenylserine			**2*S*3*S* Phenylserine**			
Protein	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	SP 3* (*S*/*R*)
wild-type eTA	92 ± 3	0.016 ± 0.002	5750 ± 740	96 ± 3	0.017 ± 0.002	5650 ± 690	1.0
Y30F R229N F87R H126R F127H	1530 ± 120	1.84 ± 0.29	829 ± 147	1500 ± 100	1.03 ± 0.16	1450 ± 240	1.8 (*S*)

Table 4.8: Kinetic parameters for the eTA catalysed cleavage of 2*S*3*R* and 2*S*3*S* phenylserine. The selectivity preference (SP) at the 3 position is calculated from k_{cat}/K_m ratios for both 2*S*3*R* and 2*S*3*S*

4.6 Summary

A summary of eTA variants is shown in Table 4.9. It was found that rational combinations of variants from the first round of CASTing could lead to a highly selective eTA. The best of these (Y30F R229N H83S F87Y) displaying a SP of 500 for 2*S*3*S* PS. All variants with His83 substitutions lead to reductions in k_{cat} whereas all those identified not targeting this position, displayed higher than wild-type values. This suggests His83 is an important catalytic residue, however not necessarily the base in the PS retro-aldol cleavage reaction mechanism. An iterative round of CASTing identified a variant (Y30F R229N H126R F127H) with a SP of 9.6. This variant displayed good catalytic efficiency with only a 4-fold reduction in k_{cat}/K_m with respect to wild-type eTA for 2*S*3*S* PS. This variant can be seen as one, identified in a total focused sequence space of 12^4 variants. Around 3200 individual colonies were screened for improved selectivity. This highlights the effectiveness of the the CASTing approach where the synergistic effect produced by all four mutations would otherwise be extremely difficult to obtain by single-site saturation mutagenesis.

CASTing group 4, targeting postions Y30 and R229, appeared to show the largest initial improvement in stereoselectivity for 2*S*3*S* PS. The R229N substitution appeared in close to half the purified 2*S*3*S* PS selective variants which when also paired to a Y30F mutation lead to a 3.1-fold selective eTA. A single Y30R variant also had a comparable level of selectivity. We aim to provide rational to this observed increase in Chapter 5.

Interestingly no variants with improved selectivity for 2*S*3*R* PS were identified. While the NDT codon incorporates 12 of the 20 canonical amino acids, it is possible that one of the eight not included, or a combination thereof, are required for such improvements in 2*S*3*R* PS selectivity. It may also be that the threshold levels set to identify hits in the initial screening plates was too high. The two-fold activity preference required for further analysis could have been lowered to allow more variants through with potentially smaller increases in 2*S*3*R* PS selectivity. We could speculate that it may only be in iterative cycles of CASTing where the larger increases in selectivity would become apparent. Another possibility is that the conditions of the screen were not suitable for identifying improved 2*S*3*R* PS selective variants. If, with respect to the wild-type enzyme, improvements in selectivity are seen by further reductions in K_m, this would be undetectable by the screen.

In order to observe measurable activity, screening conditions required 0.15 mM substrate. However, for reaction rates to be roughly proportional to k_{cat}/K_m then screening at sub-K_m concentrations is required. Otherwise, only changes in k_{cat} will be identified.

	2*S3R* Phenylserine			2*S3S* Phenylserine			
Protein	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)	SP 3* (*S/R*)
wild-type eTA	92 ± 3	0.016 ± 0.002	5750 ± 740	96 ± 3	0.017 ± 0.002	5650 ± 690	1.0
CASTing							
H83S F87Y (4 G10)	61 ± 5	0.8 ± 0.18	76 ± 18	75 ± 4	0.38 ± 0.07	197 ± 38	2.6 (*S*)
H83C F87Y (5 F9)	51 ± 3	0.65 ± 0.09	78±12	58 ± 2	0.5 ± 0.05	116±12	1.5 (*S*)
Y30F R229N (4 B10)	172 ± 4	0.66 ± 0.03	261 ± 13	412 ± 23	0.51 ± 0.07	808 ± 120	3.1 (*S*)
Y30R (3 F8)	212 ± 15	0.79 ± 0.15	268 ± 54	209 ± 8	0.25 ± 0.03	836 ± 105	3.1 (*S*)
Rational combinations							
Y30F R229N H83S F87Y	1.1 ± 0.3	10.0 ± 4.8	0.11 ± 0.06	17 ± 1	0.31 ± 0.08	55 ± 15	500 (*S*)
Y30F R229N H83S	18 ± 0.4	1.07 ± 0.12	17 ± 2	50 ± 2	0.88 ± 0.1	57 ± 6	3.4 (*S*)
Y30F R229N F87Y	48 ± 2	2.04 ± 0.15	24 ± 2	373 ± 14	0.88 ± 0.09	424 ± 46	18 (*S*)
Iterative CASTing							
Y30F R229N F87S (1 G3)	156 ± 7	0.85 ± 0.12	184 ± 27	587 ± 8	0.54 ± 0.02	1090 ± 40	5.9 (*S*)
Y30F R229N F87N (3 A10)	155 ± 9	0.85 ± 0.13	182 ± 30	650 ± 38	0.64 ± 0.11	1020 ± 180	5.6 (*S*)
Y30F R229N F87R (5 A8)	119 ± 12	1.05 ± 0.28	113 ± 32	414 ± 13	0.55 ± 0.05	752 ± 72	6.7 (*S*)
Y30F R229N H126R F127H (2 B7)	368 ± 24	2.81 ± 0.41	131 ± 21	555 ± 25	0.44 ± 0.07	1260 ± 210	9.6 (*S*)
Y30F R229N F127I (4 A12)	405 ± 20	2.80 ± 0.31	144 ± 18	560 ± 31	1.17 ± 0.19	478 ± 82	3.3 (*S*)
Y30F R229N H126R F127L (6 B1)	371 ± 24	2.24 ± 0.34	166 ± 27	507 ± 28	0.93 ± 0.16	545 ± 99	3.3 (*S*)
Y30F R229N F127C (6 E12)	455 ± 33	4.69 ± 0.67	97 ± 16	598 ± 28	2.99 ± 0.30	200 ± 22	2.1 (*S*)
Iterative CASTing rational combinations							
Y30F R229N H126R	1770 ± 250	2.53 ± 0.63	699 ± 199	2560 ± 170	2.38 ± 0.26	1080 ± 140	1.5 (*S*)
Y30F R229N F127H	1860 ± 150	2.51 ± 0.36	740 ± 122	1960 ± 90	0.82 ± 0.10	2390 ± 310	3.2 (*S*)
Y30F R229N F87R H126R F127H	1530 ± 120	1.84 ± 0.29	829 ± 147	1500 ± 100	1.03 ± 0.16	1450 ± 240	1.8 (*S*)

Table 4.9: Kinetic parameters for the purified variants from CASTing and ISM. Kinetics measured for the enzyme catalysed cleavage of 2*S*3*R* and 2*S*3*S* PS. The selectivity preference (SP) at the 3 position is calculated from k_{cat}/K_m ratios for both epimers.

Chapter 5

Structural Insights into *E. coli* Threonine Aldolase Stereoselectivity

5.1 Structural guidance in enzyme engineering

There are now over 130,000 protein structures in the protein data bank (PDB). If we were to look back 20 years this number falls short of 9,000[1]. Advances in structural biology are thanked for this sharp rise. X-ray crystallography, the original technique for obtaining these macromolecular structures, is now one of several methods used. Neutron crystallography, electron microscopy (EM) and nuclear magnetic resonance (NMR) spectroscopy all provide their own unique approaches facilitating the growth of the PDB. With atomic resolution structures it is possible to observe interactions such as how an inhibitor binds to a receptor [Christen et al., 2012] or how misfolded proteins interact in aggregation [Moreno-Gonzalez and Soto, 2011].

One area that has benefited from advances in structural biology is protein engineering. Directed evolution on enzymes with unknown structures will likely require random mutagenesis across a gene. This in turn can require huge screening efforts before desirable traits are identified. However, a well defined structure will localise engineering efforts to an active site, where mutations leading to improved enantioselectivity [Reuben et al., 2008], substrate specificity [Wells et al., 1987] and new activities [Williams et al., 2005] are often located.

[1] Statistics from https://www.rcsb.org/stats/growth/overall

The methods listed for obtaining high resolution structures only offer ways of obtaining static snapshots of proteins at a particular moment in time. Computational modelling in enzymology offers an *in silico* approach to visualising these structures at work over picosecond (psec) to microsecond time scales. This is only possible, however, due to the wealth of structural information we have on proteins, their functions, understanding of laws governing atomic and sub-atomic interactions and increases in computational power.

The aims of this chapter are two-fold. Firstly, to use X-ray crystallography to obtain a high resolution structure of *E. coli* threonine aldolase (eTA). Secondly we will use this structure for computational modelling to try and understand the poor stereoselectivity of the enzyme with phenylserine (PS), which may in turn lead to an understanding of the effects of the mutations described in Chapter 4.

5.2 A high resolution structure of *E. coli* threonine aldolase

The first structure of a L-threonine aldolase (TA) was published in 2002 from *Thermotoga maritima* [Kielkopf and Burley, 2002]. Since then multiple L-TA structures have been solved. More recently the first structure of a D-TA, from *Alcaligenes xylosoxidans* was obtained [Uhl et al., 2015] revealing an almost mirror symmetry of the L-theonine aldolase active site residues.

di Salvo et al. published several structures of a *low-specificity* L-TA from *E. coli* bound to threonine, glycine and serine at 2.1 Å resolution (PDB codes 4LNL, 4LNM and 4LNM) [di Salvo et al., 2014]. A sequence alignment of that gene revealed a 98.8% sequence identity with the eTA used in this project, highlighting four amino acid changes. It was therefore decided that an improved structure of the eTA from this project should be obtained.

5.2.1 Crystallisation

Crystallisation conditions for an *E. coli* TA had previously been determined [di Salvo et al., 2014]. These showed that diffracting enzyme crystals could be grown in 0.2 M

$MgCl_2$ hexahydrate, 0.1 M HEPES (pH 7.5), 30% PEG 400 with 600 μM of the enzyme. eTA was purified using nickel affinity chromatography (Section 2.4.1) and subject to size exclusion chromatography (Section 2.4.6). Eluted fractions were analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (Section 2.4.5) with those containing eTA being pooled and concentrated to 9 mg mL^{-1} (Section 2.4.6). Crystal trays were set up containing 0.6 mM eTA, 0.2 M $MgCl_2$ hexahydrate, 0.1 M HEPES (pH 6.8-8.0) and PEG 400 (20-34% (v/v)). The sitting drop vapour diffusion method was performed for crystal growth which produced large single crystals across multiple conditions after 8 days at 20 ^{o}C. One particularly good crystal-forming condition is: 0.2 M $MgCl_2$ hexahydrate, 0.1 M HEPES pH 7.6, 28.9% (v/v) PEG 400 with 0.6 mM eTA (Figure 5.1). It was from this drop that the final refined structure is produced. Crystals were selected and flash cooled in liquid nitrogen prior to data collection at Diamond Light Source synchrotron facility (Oxford, UK).

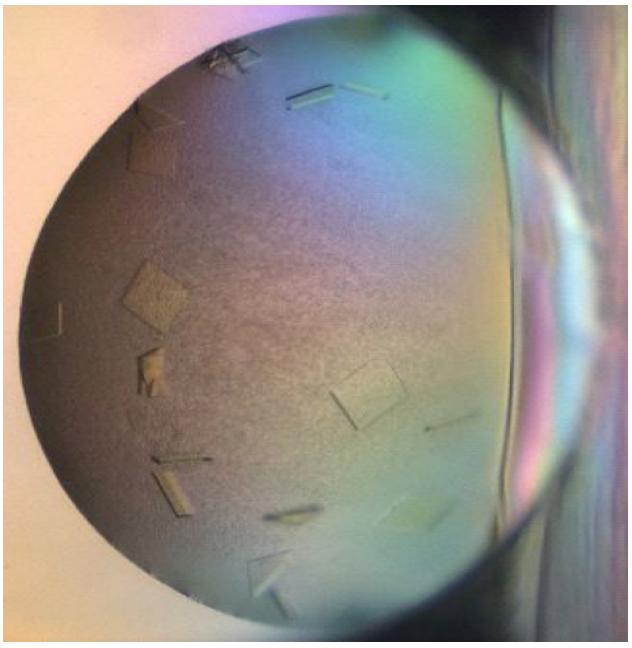


Figure 5.1: Image of drop showing crystal formation after 8 days in 0.2 M $MgCl_2$ hexahydrate, 0.1 M HEPES pH 7.6 and 28.9% (v/v) PEG 400.

5.2.2 Structure determination of eTA

Data sets for crystals were collected at Diamond Light Source on the macromolecular crystallography beamline I02 at 100K. The diffraction images were integrated, scaled and reduced using a set of programs implemented as part of the Xia2 program suite [Winter,

2010]. Five percent of the reflections from the collected datasets were selected at random and excluded from the refinement and constituted the R_{free} set. Two comparable data sets to 1.6 Å resolution were obtained and efforts were focused on a single one of these. The dataset belong to space group $P222_1$ with unit-cell parameters $a = 77.2$ Å, $b = 101.5$ Å, $c = 175.9$ Å. There are two eTA molecules (present as a dimer) in the asymetric unit cell with a solvent content of 47%. The crystal structure was determined by molecular replacement using the program PHASER [McCoy et al., 2007] with the similar *E. coli* TA structure (PDB code 4LNL) [di Salvo et al., 2014] as a search model. One round of rigid body refinement followed by multiple rounds of restrained refinement were performed using REFMAC [Murshudov et al., 1997] with rebuilding into both $2F_{obs}$-F_{cal} and F_{obs}-F_{cal} maps using COOT [Emsley et al., 2010]. Data statistics for the refined eTA are shown in Table 5.1.

The threonine aldoalse family are reported to function as tetramers [Liu et al., 1997, Kielkopf and Burley, 2002, Qin et al., 2014]. The asymmetric unit of the crystal contains two identical protamers (monomers) of eTA with the homotetramer generated around a crystallographic two-fold symmetry axis (Figure 5.2). Those protomers from the asymmetric unit cell are shown as chains A and D, where chains B and C are the symmetry generated mates (Figure 5.2A). Six magnesium ions are observed in the tetramer, two of these position closely at the interface between all four chains where it is proposed they play a role in oligomerisation [di Salvo et al., 2014]. Two water molecules coordinate between these ions and further interactions from the negatively charged oxygen atoms of Ala93 (backbone) Val94 (backbone) and Ser97 (sidechain) from all four monomers are observed with the magnesium ions (Figure 5.2B). Two equivalent active sites are observed in the asymmetric unit creating four equivalent active sites in the functional tetramer. Interestingly residues from three chains form each active site (Figures 5.2C and D). Active site 1 is composed primarily of chain A with His83, Phe87, Lys197 and Arg308 where the PLP cofactor is found as a Schiff base to Lys197. Tyr30, Lys222 and Arg229 from chain D and His126 from chain B are also apparent within this site.

Covalently bound PLP is observed in both active sites of the asymmetric unit (Figures 5.3A and B). Interestingly the density of the covalently bound PLP-Lys197 suggested a far from planar orientation for the imine Schiff base (N_{SB}) and the pyridinium ring. A dihedral measurement between C3-C4-C4*-N_{SB} atoms highlights this non-coplanarity (see

	eTA
Diamond beamline	I02
Space group	$P222_1$
Cell dimensions a, b, c (Å)	77.2, 101.5, 175.9
α, β, γ (^{o})	90, 90, 90
Resolution range (Å) *	58.62-1.60 (1.64-1.60)
R^*_{merge}	0.075 (0.852)
R_{pim} (I)*	0.029 (0.460)
R_{pim} (I +/-)*	0.038 (0.596)
Observed reflections	635899
Unique reflections	90793
Completeness (%)*	99.7 (96.6)
Multiplicity*	7.0 (5.0)
$<I/\sigma(\sigma)^*$	14.6 (1.7)
$CC^*_{1/2}$	0.998 (0.512)
Refinement	
R_{factor} (%)	0.159
R_{free} (%)	0.187
No. of protein atoms	5378
No. of solvent molecules	265
Average overall B-factor) (Å^2)	27.5
RMS bond length (Å)	0.020
RMS bond angles (^{o})	2.0
Ramachandran analysis	
Most favoured	98.2
Outliers	0

$R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - < I(hkl) > | / \sum_{hkl} \sum I_i(hkl)$
* Values in parentheses are for highest-resolution shell

Table 5.1: X-ray crystallographic data collection and refinement for eTA. Values given in parentheses correspond to the outermost shell of the resolution range. Ramachandran analysis performed using MolProbity [Davis et al., 2007].

nomenclature in Figure 5.3C). These angles were measured as 65^o and 56^o in active site 1 and 2 respectively. The 9^o difference is likely due to the decreased density for Lys197 in chain A. This non-coplanarity has not been mentioned in any TA literature where similar cases are observed. These issues have however been addressed for other PLP-fold type I enzymes [Dajnowicz et al., 2017, Hayashi et al., 2003, 1998]. Aspartate amino transferases also rely on a PLP dependent mechanism. Structures show when covalently bound to the enzyme the PLP N_{SB}-C3* non-coplanarity is observed [Hayashi et al., 1998]. It is proposed that a significant increase in the imine N_{SB} pK_a occurs during the transaldimation whereby

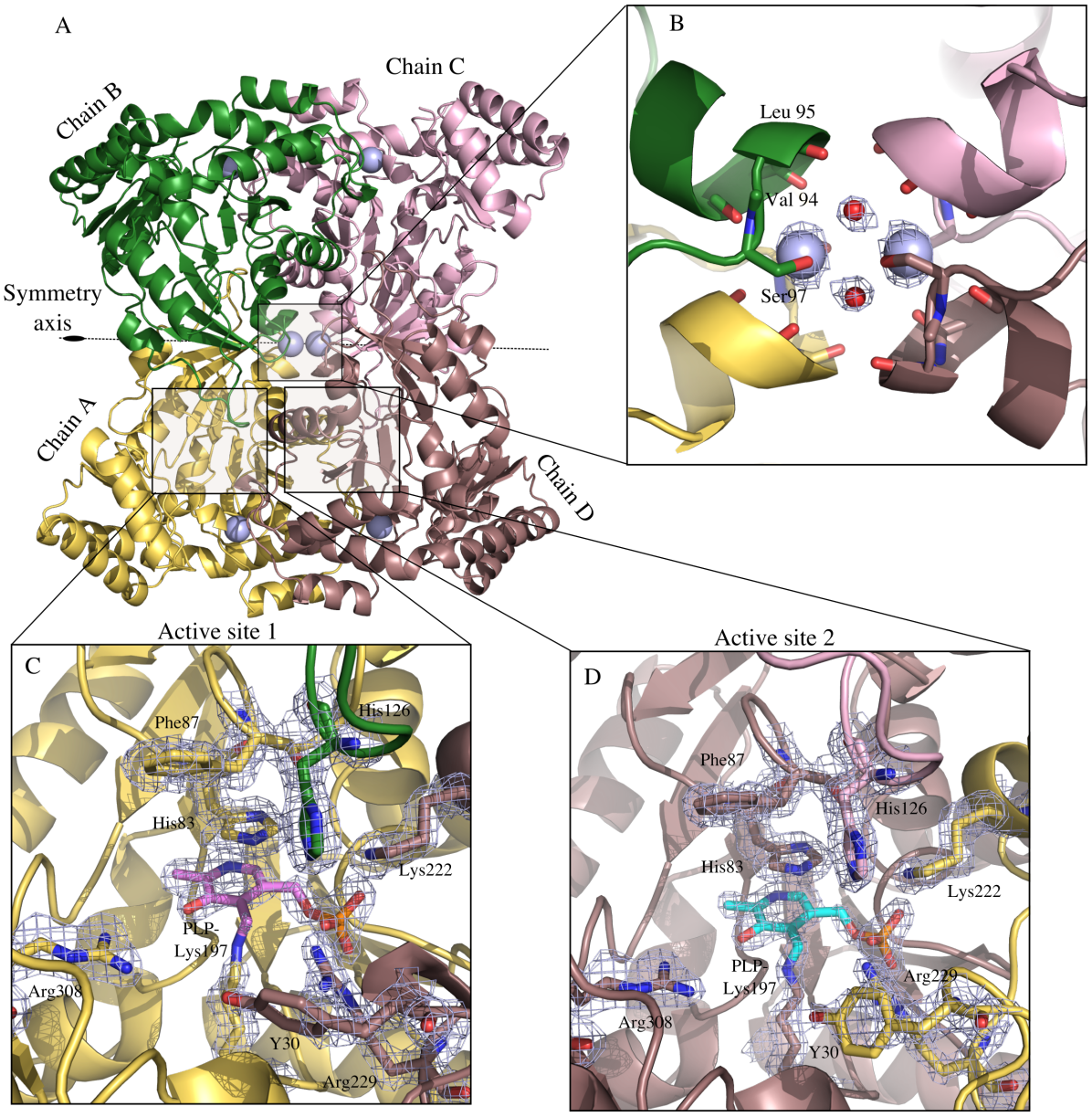


Figure 5.2: A: The tetrameric structure of eTA generated from two protamers of the asymmetric unit dimer along a crystallographic symmetry axis. The four protein chains are labelled A (yellow) B (green) C (pink) and D (brown). B: The interface between the four chains highlights two magnesium ions that coordinate with the negatively charge oxygen atoms of adjacent residues in each monomer. The $2F_{obs}$-F_{cal} map of these metal ions and two closely coordinated waters is shown at at one sigma level. C: Active site 1 predominantly formed of chain A where residues from chains B and D also identify within this site. The PLP ligand (pink) is clearly seen in the covalently bound form to Lys197 of chain A. The $2F_{obs}$-F_{cal} map is shown at one sigma level for key active site residues and the PLP cofactor. D: Active site 2 of the asymmetric unit dimer whereby chains A, C and D contribute residues. The PLP ligand is observed (blue) covalently bound via a Schiff base to Lys197 of chain D. The $2F_{obs}$-F_{cal} map is shown at one sigma level.

lysine is released and substrate is bound (Figure 1.17). As a result it is likely that two different protonation states occur at N_{SB} in the enzyme bound and ligand bound forms

of PLP. Protonation of N_{SB} in the ligand bound form would likely produce a stabilising hydrogen bonding interaction with the PLP oxygen (O3*) and create N_{SB} and pyridine ring coplanarity. This is consistent with structures showing the ligand bound forms of PLP with threonine [di Salvo et al., 2014] and glycine [Qin et al., 2014]. To investigate whether the N_{SB} torsion is an effect of the surrounding environment, Dajnowicz et al. [Dajnowicz et al., 2017] performed density function theory (DFT) calculations on a truncated PLP-Lys ligand altering the protonation between the N_{SB} and O3*. Interestingly this showed that when both of these positions are unprotonated, the C3-C4-C4*-N_{SB} dihedral angle is 44^o. If either the N_{SB} or O3* are protonated then the torsion is removed and the N_{SB} becomes coplanar to the pyridinum ring.

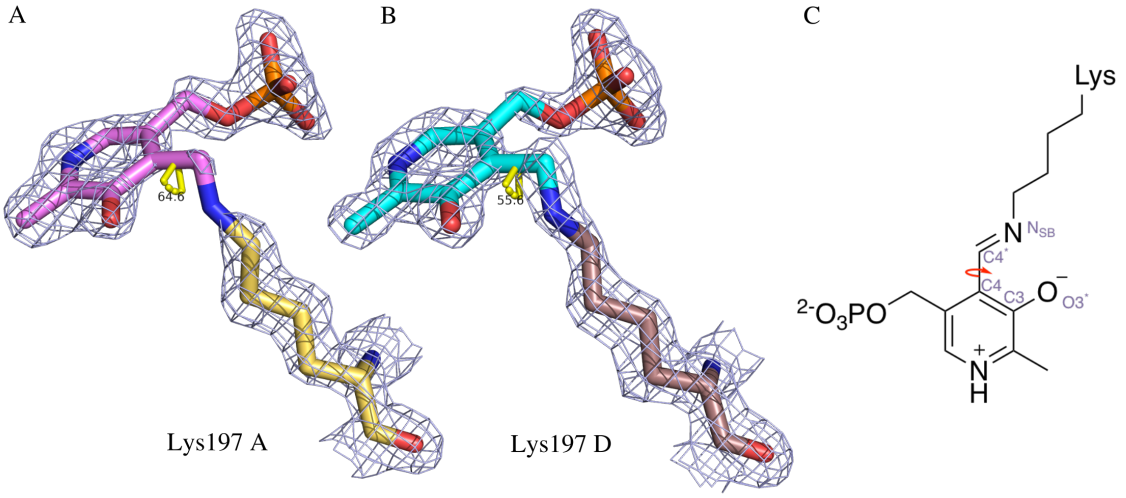


Figure 5.3: A: The PLP Schiff base with Lys197 in chain A. A 65^o torsion is observed between the N_{SB} and the pyridinium ring. The $2F_{obs}$-F_{cal} map is shown at one sigma level. B. PLP covalently linked to Lys197 by a Schiff base in chain D. More continuous density is observed around the N_{SB} where a torsion angle of 56^o is observed to the pyridinium ring. C: A schematic of the PLP-Lysine highlighting the likely protonation state of the N_{SB} and the O3* based on DFT calculations [Dajnowicz et al., 2017]. The nomenclature of atoms used in measuring the torsion angle between the N_{SB} and the pyridinium ring are shown in indigo.

5.3 Structural modelling of *E. coli* threonine aldolase

Structural modelling of enzymes through computational (*in silico*) approaches is an ever increasingly popular method of studying enzyme systems [van der Kamp and Mulholland, 2013]. Quantum mechanics/molecular mechanics (QM/MM) molecular dynamics (MD) simulations have been effective in elucidating the mechanisms of enzymes [Bowman et al., 2008, Daniels et al., 2014]. They use QM methods (semi empirical, DFT and *ab initio*) to study the electron rearrangements, involved in bond making and breaking, while simultaneously sampling the surrounding environment by simpler MM methods. With relatively small QM regions, semi empirical QM/MM calculations are generally less computationally expensive than those using DFT or molecular orbital *ab initio* QM approaches alone [Siegbahn and Himo, 2009]. However, they are more expensive than those simulations solely reliant on MM [Warshel, 2003]. No single approach can be considered as the correct way of studying enzyme reactions. Successful methods in similar systems would however suggest a sensible starting point when choosing the modelling approach. QM/MM has been used to observe the enantioselective proton transfer in the mechanism of citrate synthase for the conversion of fluroacetyl-CoA to flurocitrate [van der Kamp et al., 2011]. Similar QM/MM methods have also been used to study the stereoselectivity of *N*-acetylneuraminic acid lyase; an aldolase catalysing the reversible condensation of pyruvate and *N*-acetyl-D-mannosamine [Daniels et al., 2014].

MM and QM/MM methods are used for modelling of the high resolution eTA crystal structure obtained in Section 5.2. Using these approaches we aim to identify how the stereochemical step of the reaction proceeds.

5.3.1 Choosing the stereoseletive reaction step

The stereoselectivity of eTA is determined by the free energy difference between the transition state ($TS^{\ddagger}$) barriers for each of the possible diastereomers in the aldol/retro-aldol step of the reaction[2] (Figure 5.4A). That is, the proton transfer or abstraction at the aldol oxygen (Figure 5.4B). It is not possible to model the the whole eTA catalysed

[2]Assuming the reaction is under kinetic control

mechanism, but we are concerned with how the enzyme (particularly the variants described in Chapter 4) can achieve stereoselectivity. We therefore concentrate our modelling on the stereogenic step in the aldol condensation of glycine and benzaldehyde (Figure 5.4B). As the 2*S* stereochemistry remains fixed at the α-carbon, the resulting phenylserine (PS) diastereomeric products will be referred to by their 3*R* or 3*S* stereochemistry.

5.3.2 Modelling of the wild-type eTA

The PLP-glycine ligand (PXG) shown in Figure 5.4 was built for *in silico* modelling. The protonation state of PXG was determined whereby the N_{SB} is protonated. This is in agreement with that described by Fesko et al. in *Aeromonas jandaei* TA [Fesko et al., 2018]. Additionally, as discussed in Section 5.2, studies on other PLP enzymes have shown the pK_a of the PLP imine group to be around 8.8 when in a non-enzyme bound Michaelis complex [Hayashi et al., 1998, 2003]. The most suitable pre-simulation orientation of the PXG ligand was determined by the previously solved *E. coli* TA structure (PDB code 4LNL) with the PLP-threonine ligand bound. With structurally aligned protein chains, PXG could be manually docked into a ligand bound crystallographic orientation. The unbound Lys197 could also be oriented into a suitable rotamer (Figure 5.5A).

Two benzaldehyde molecules were also separately manually docked into the eTA active sites (Figure 5.5B). A simple flip of the carbonyl group created two orientations that could lead to *si* or *re* face attack, in turn creating the 3*R*, or 3*S*, products respectively. From here, the benzaldehyde orientations leading to the formation of either 3*R*, or 3*S*, products (Figure 5.4) will be referred to as pre-3*R*, or pre-3*S* structures. Parallel simulations of eTA with pre-3*R*, or pre-3*S*, benzaldehyde orientations are performed for comparable analysis.

Enlighten is a set of protocols and tools to run (automated) atomistic simulations of enzyme-ligand systems [3]. It was created by Dr Marc Van der Kamp (University of Bristol, UK) and utilises the Assisted Model Building with Energy Refinement (AMBER) suite of programs (AmberTools). *Enlighten* provides a logical, stepwise progression through simulations that can be broken down into several parts. The preparatory step (PREP) generates ligand parameters, a solvent sphere, adds hydrogens and generates AMBER

[3] *Enlighten* is free and available for download from https://github.com/marcvanderkamp/enlighten

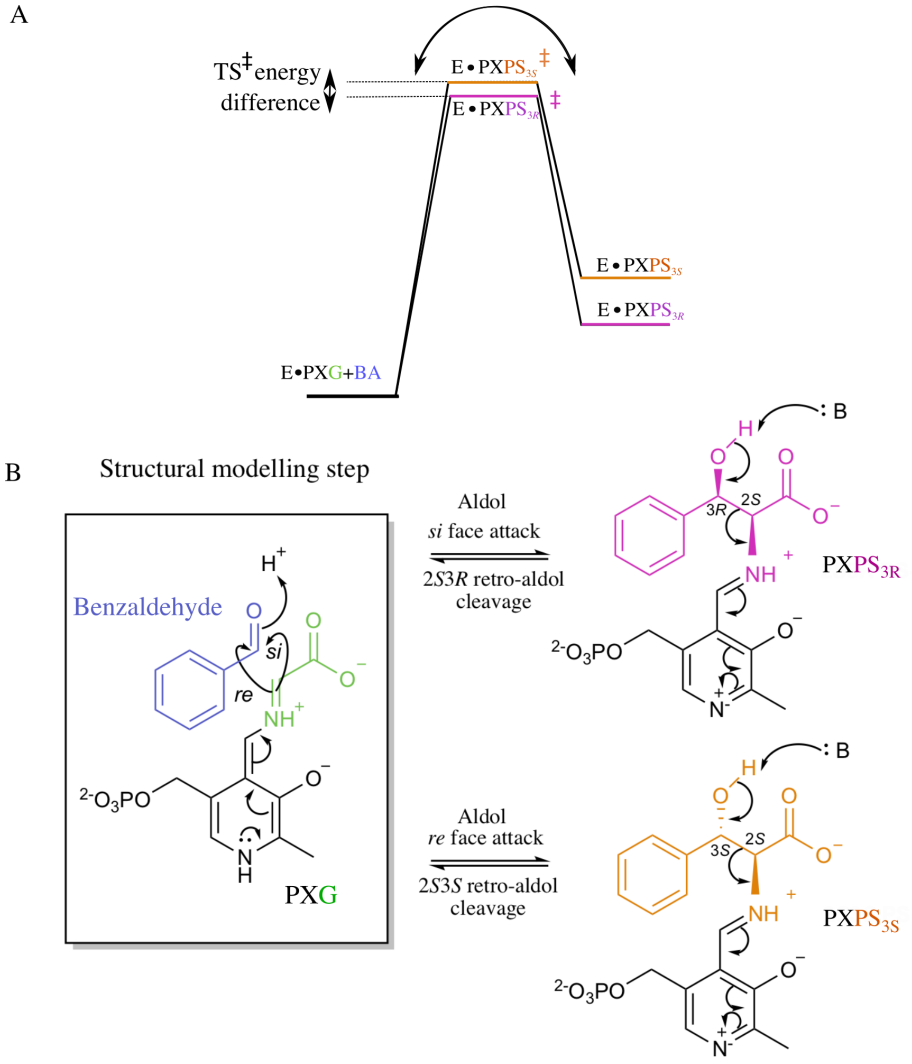


Figure 5.4: A: Free energy diagram for the proposed eTA stereoselectivity determining step. The difference in transition state (TS) energies for the 3*R* and 3*S* PS products in the C-C bond forming step is expected to be the stereoselectivity determining factor in the reaction. The greater the difference in this barrier the greater the expected difference in stereoselectivity. B: Proposed mechanisms for the aldol and retro-aldol steps of PS synthesis and cleavage. The structural modelling step is highlighted whereby the C-C bond formation between benzaldehyde (blue) and the PXG ligand (glycine (green) as a Schiff base with PLP (black)) is mediated by transfer of a proton onto the benzaldehyde oxygen. The additional electron density, delocalised in the PLP pyridine N, can then move back along the conjugated π bond system allowing for the Schiff base imine to attack onto either the *re* or *si* face of benzaldehyde determining the C3 stereochemistry. The retro-aldol cleavage of the $PXPS_{3R/3S}$ ligand (2*S*3*R* PS (pink) or 2*S*3*S* PS (orange) with a Schiff base to PLP (black)) is shown on the right of the equilibrium arrow whereby a catalytic base abstracts the hydroxyl proton and the C-C bond is cleaved.

topology and coordinate files. The structural optimisation step (STRUCT) takes the generated topology and coordinate files and performs a brief simulated-annealing and minimisation protocol. The now-optimised structures can be passed into MD simulations (DYNAM).

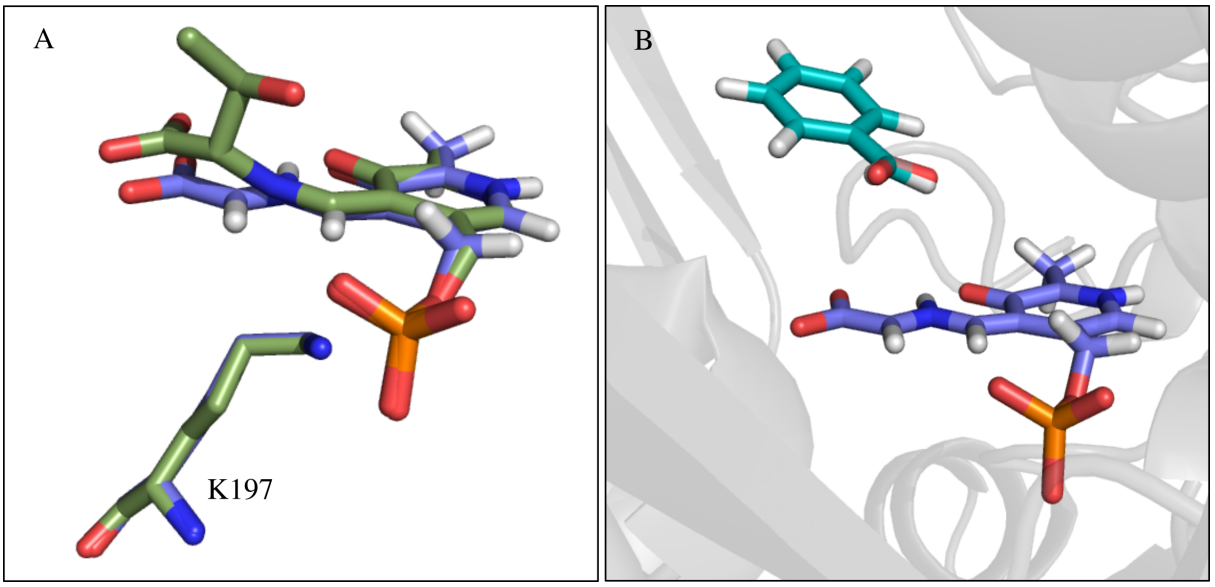


Figure 5.5: Panel A: The PXG ligand for modelling (indigo) manually docked into position with respect to the *E. coli* TA PLP-threonine ligand (green) (PDB structure, 4LN)[di Salvo et al., 2014]). The Active site lysine (K197) has also been moved into a similar rotamer as that of the ligand bound structure. Panel B: Pre-3*S* (brown) and pre-3*R* (blue) benzaldehyde were manually docked into a suitable pre-simulation position relative to the PXG ligand (indigo). A simple flip of the benzaldehyde ligand was performed for either pre-3*R* or pre-3*S* orientations. Parallel simulations for each of these are performed.

The *Enlighten* protocol was performed on eTA with PXG and benzaldehyde (in the pre-3*R* and pre-3*S* orientations)(Figure 5.6). From PREP, a solvent sphere of 20 Å around the ligands was formed in addition to existing crystallographic waters. The pK_a values for sidechains within the sphere were predicted using propka [Søndergaard et al., 2011] and the protonation states assigned. STRUCT performed a brief energy minimisation, optimising the structure for further MD simulations. For both benzaldehyde orientations an obvious stabilising salt bridge formed between the carboxylate of PXG and Arg308. The manually docked benzaldehyde in a pre-3*R* orientation remained in a similar position after the minimisation suggesting this to be an energetically favourable pose (Figure 5.6C). When in a pre-3*S* orientation the energy minimisation showed a more distinct rotation of the ligand away from the initially docked pose (Figure 5.6D).

5.3.3 MM MD simulations

The minimised structures offer a static snapshot of two possible ligand conformations. In reality these ligands will not remain in a fixed position and instead are free to move. Using

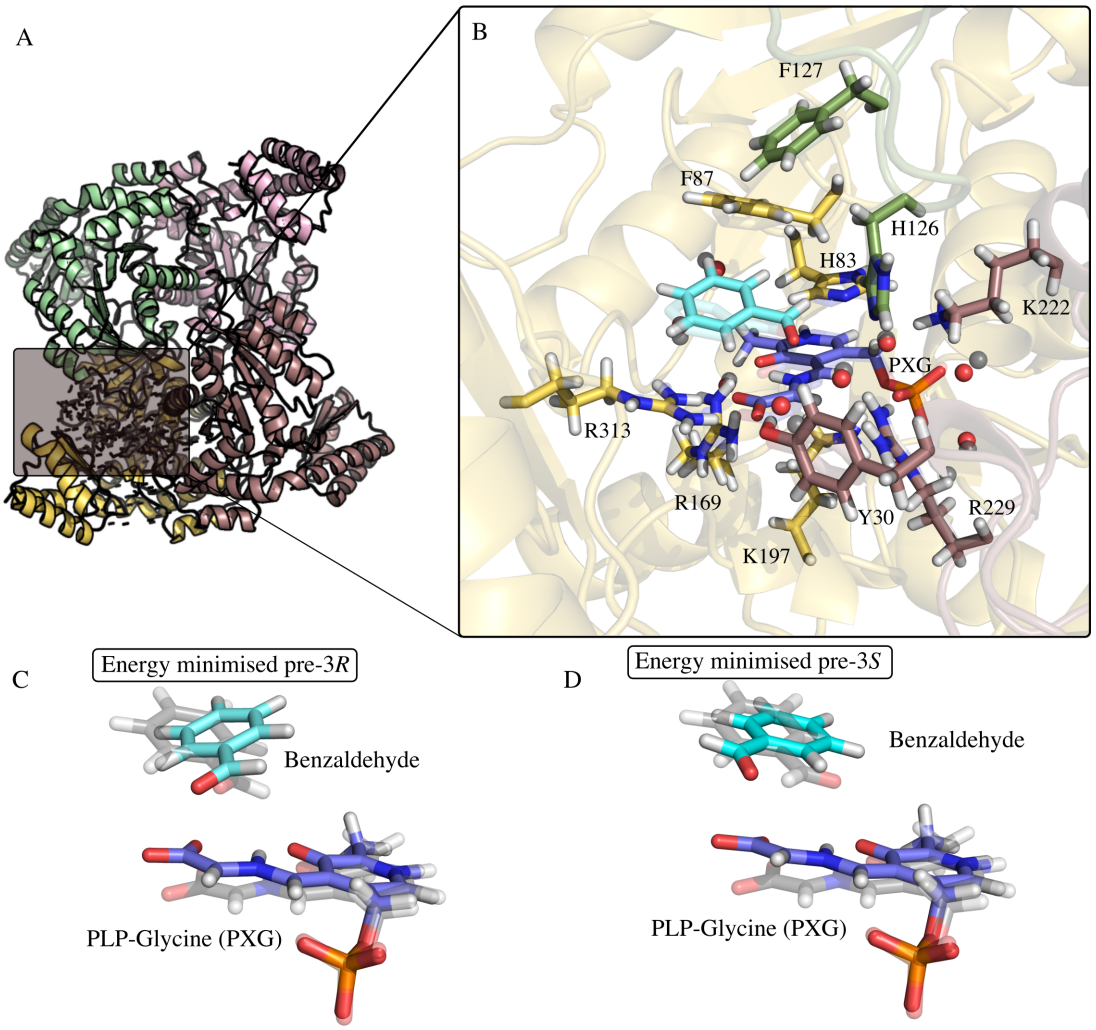


Figure 5.6: A: The eTA homotetramer coloured by each monomer chain. A 20 Å solvent sphere is shown behind the expansion panel. The homotetramer is coloured by chain: chain A (yellow), chain B (green) chain C (pink) and chain D (brown) . B: Energy minimisation from STRUCT for benzaldehyde in a pre-3*R* orientation. Residues and ligands following the STRUCT minimisation are shown as opaque and those prior to the simulation are shown as transparent. Residues within close proximity of the ligands are shown and coloured according to their chain. Water molecules within 3 Å of the ligands are are shown before and after STRUCT as red and grey spheres respectively. C: The energy minimised structures of benzaldehyde (light blue) in a pre-3*R* orientation and PXG (indigo) from STRUCT. The pre minimised ligands are shown as transparent for reference. D: As C but for benzaldehyde in a pre-3*S* orientation.

MM MD simulations we aim to observe how the ligands behave within the active site. Furthermore we can used these dynamic simulations to predict the likelihood of a reactive Michaelis complex forming in the pre-3*S* and -3*R* orientations. To ensure the benzaldehyde ligand did not move away from the active site during the initial heating phase, a restraint was applied (Section 2.7.2). For this a reaction coordinate is defined between the carbonyl carbon of benzaldehyde (CA) and the α-carbon of PXG (CG2) where the aldol C-C bond is

expected to form. This coordinate is subsequently referred to as d(C-C) (Figure 5.7.A)

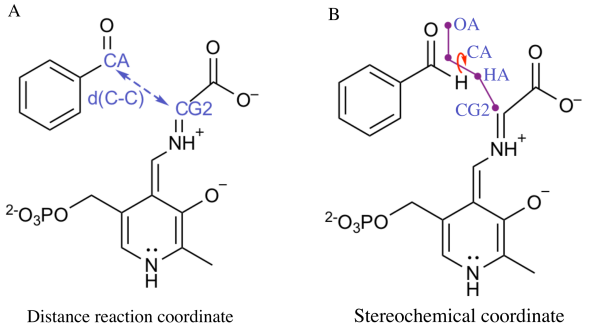


Figure 5.7: A: The defined distance reaction coordinate (d(C-C)) between the C-C bond forming atoms of benzaldehyde (CA) and PXG (CG2). B: Stereochemical coordinate measuring the dihedral between the benzaldehyde oxygen (OA), carbonyl carbon (CA) and proton (HA) and the α-carbon of PXG (CG2).

20 separate MD simulations (3*R*/3*S* MD1-20), for each of the benzaldehyde orientations were performed. Each simulation was 150 psec, where the first 50 psec included an initial heating phase to 300 K and a distance restraint along the reaction coordinate (d(C-C)) of 4 Å was applied. This restraint was then removed for the final 100 psec of unrestrained simulation. The distances along the d(C-C) reaction coordinate were plotted, as shown in Figure 5.8. It is clear to see that when the distance restraint is removed there is a far greater range of movement between the ligands in both sets of simulations. A stereochemical coordinate is also defined to investigate the possibility of each of the pre-3*R* or -3*S* ligands "flipping" into the opposite orientation. For this a dihedral between benzaldehyde and the PXG is measured (Figure 5.7B). This measured the dihedral angle between the benzaldehyde oxygen (OA), carbonyl carbon (CA) and proton (HA) and the α-carbon of PXG (CG2). The results show that throughout all MD simulations the benzaldehyde ligands remain mostly in a pre-3*R* or -3*S* conformation. Where values are close to 0, 180 or -180^o, this would suggest a planar orientation to the PXG CG2 atom and subsequently favour neither a 3*R* or 3*S* product. Those simulations displaying these planar orientations only occur very briefly (for a few psec) so shall not be considered as significant with respect to switching orientation.

To obtain representative structures for the conformations sampled in simulation, clustering analysis is performed [Shao et al., 2007]. This allows for similar conformations to be partitioned into a representative structure based on the RMSD value of the cluster. Figure 5.9 shows the clustering on the RMSD of the benzaldehyde across the pre-3*R* and -3*S* MM

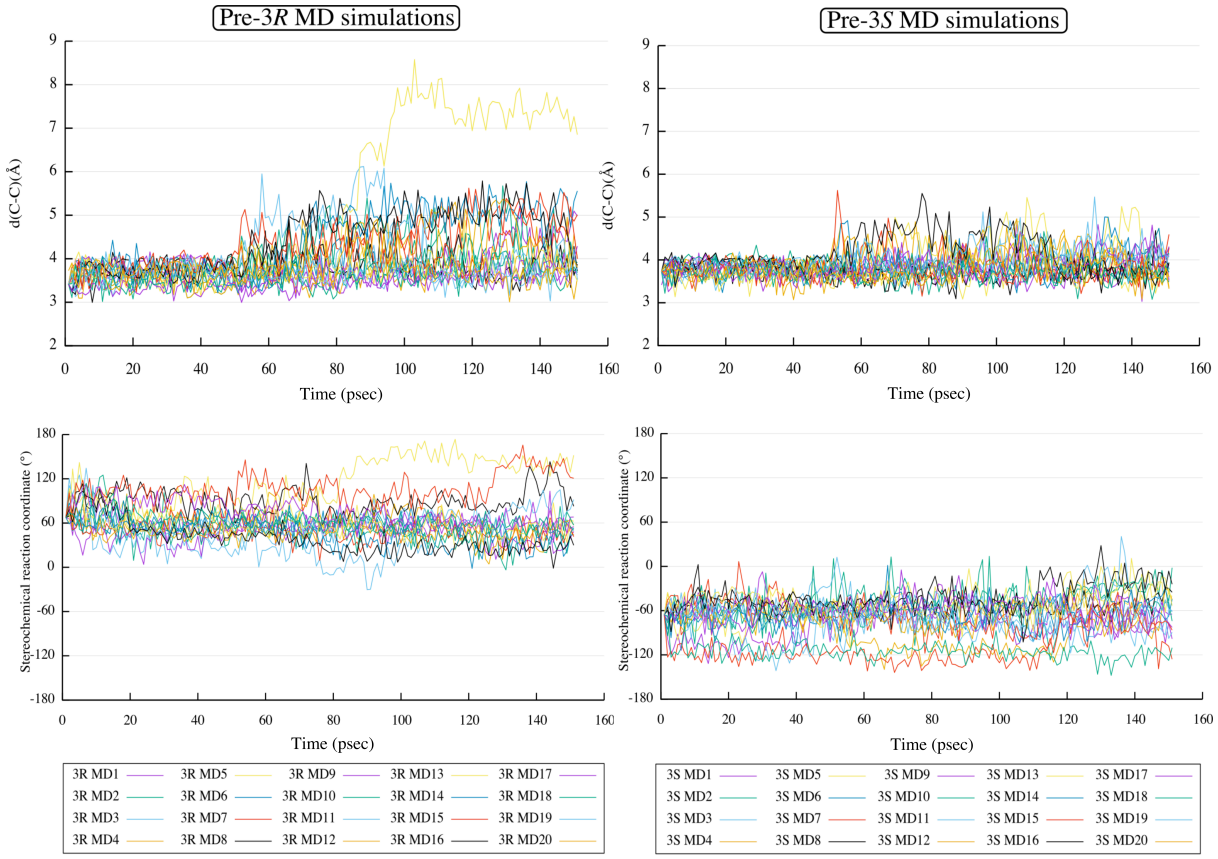


Figure 5.8: Top left: Changes in the distance reaction coordinate d(C-C) over the 150 psec simulation for pre-3*R* benzaldehyde orientations. A restraint of 4 Å was applied for the first 50 psec and then removed for the final 100 psec. Bottom left: Stereochemical coordinate against time for pre-3*R* benzaldehyde. To remain in a pre-3*R* orientation the reaction coordinate must be between 0 and 180^{o}. Top right: As top left but for pre-3*S* benzaldehyde orientations. Bottom right: As bottom left but for pre-3*S* benzaldehyde orienations. To remain in this orientation the stereochemical coordinate must be between 0 and -180 ^{o}.

MD simulations. The position of PXG is not significantly different for the two representative conformations. However, the benzaldehyde orientations are almost perpendicular to each other with their oxygen atoms in very similar positions.

To investigate how many of the MD simulations formed a potentially reactive Michaelis complex the percentage of those with a d(C-C) of less than 3.5 and 4.0 Å were measured. These were based on the final 50 psec of the unrestrained simulations (Table 5.2). The results show that roughly the same number of trajectories form a sub 3.5 Å complex with 13% of pre-3*R* and 12% of pre-3*S* simulations. This suggests that there is no preference for binding of either orientation of benzaldehyde in the wild-type eTA model. If the d(C-C) reaction coordinate distance is increased to 4 Å the results show a significantly higher proportion of trajectories within this limit. 50% are seen in a pre-3*R* and 65% in a pre-3*S*

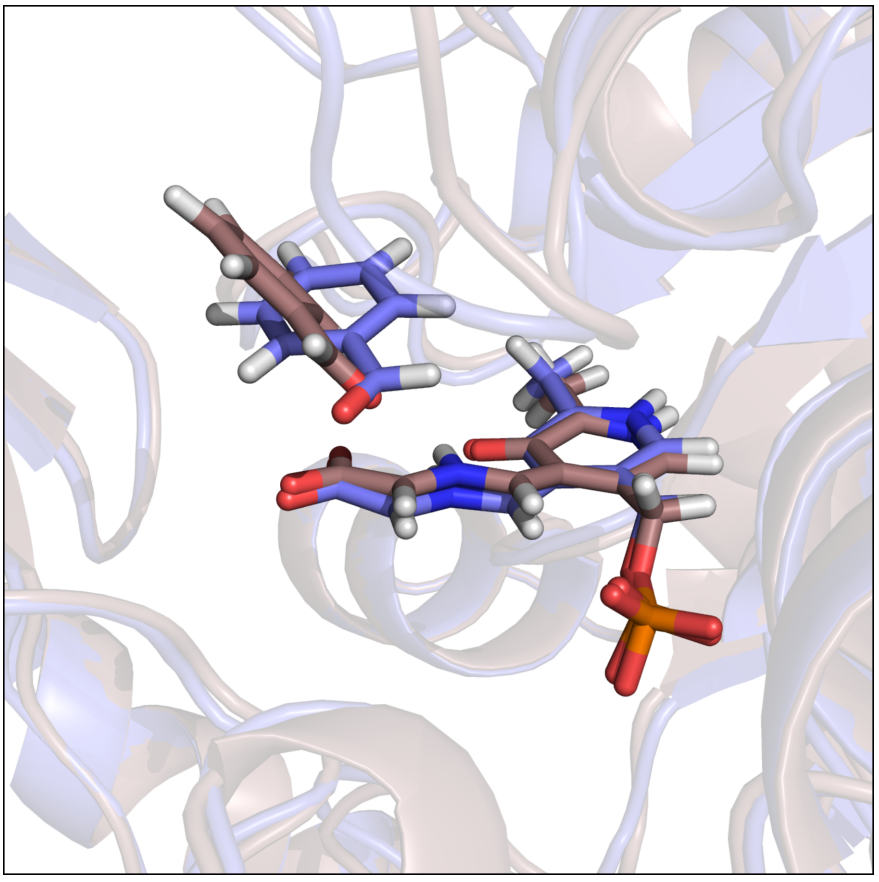


Figure 5.9: Representative conformations from MM MD simulations obtained from clustering. Benzaldehyde and PXG ligands from the pre-3*R* (purple) and pre-3*S* (brown) simulations are shown.

orientation. This suggests a small 1.3 fold 3*S* preference for binding at the increased d(C-C) distance. It is tempting to relate this binding indifference to the eTA K_m values for 2*S*3*R* and 2*S*3*S* PS in Table 4.9 as these are also virtually identical.

Enzyme / ligand orientation	d(C-C) < 3.5 Å	d(C-C) < 4.0 Å	d(C-C)< 3.5 Å 3*R*/3*S*	d(C-C)< 4.0 Å 3*R*/3*S*
eTA pre-3*R*	13%	50%	1.08 (*R*)	1.28 (*S*)
eTA pre-3*S*	12%	64%		

Table 5.2: Percentage of reactive poses across the last 50 psec of 20 MM MD simulations in eTA for pre-3*R* and pre-3*S* benzaldehyde orientations. Reactive poses are classed as those with a d(C-C) of < 3.5 Å or < 4 Å. The same percentage of simulated poses reach a sub 3.5 Å complex in both pre-3*R* and pre-3*S* simulations. Increasing the distance to 4 Å suggests that a pre-3*S* complex is preferred. The ratio of these at each distance is given to highlight any preferential reactive binding.

5.3.4 QM/MM modelling of eTA

While the unrestrained MM MD simulations may give an idea of preferential binding for one orientation over another, we are unable to simulate the C-C bond forming step with this approach. Through QM/MM modelling we will use semi-empirical calculations to treat PXG and benzaldehyde ligands with QM methods and their surroundings with MM methods (Section 2.7.3). This in turn can provide an in depth analysis of the electrostatic interactions between the ligands and their surroundings while in a reactive conformation. Furthermore, QM/MM MD simulations can be used to obtain free energy profiles of reactions. This is useful when identifying differences in free energy between *in silico* transition states (TS).

The QM/MM simulations are performed using the umbrella sampling (US) method [Kumar et al., 1992] (Section 2.7.3). This involves sampling as distance decreases along the reaction coordinate d(C-C) between benzaldehyde and PXG. The distance is reduced in 0.1 Å increments until less than that of an expected C-C bond (1.6 Å). Different trajectories from the previous unrestrained MM MD simulations, forming reactive Michaelis complexes, are used as the starting structures for each of the umbrella sampling simulations. We might therefore expect to see multiple plausible transition state structures. The umbrella sampling simulations will aim to identify key interactions in the C-C bond forming step. Furthermore energy profiling of transition state barriers relating to the formation of the 3*R* and 3*S* products is used as a more quantitative measure of stereoselectivity [Kumar et al., 1992]. We could expect these to be representative of the TS barriers for C-C bond formation which are used as a definable measure of stereoselectivity (see Section 5.3.1).

Three representative QM/MM umbrella sampling simulations for each of the product forming stereochemistries were obtained. These were all started from separate MM MD trajectories as shown in Table 5.3. Although the proton transfer in the modelling step was not expected to occur, an interesting case is observed when a 3*S* MD12 starting trajectory was used. This will be discussed further in Section 5.3.4.2.

To avoid structural fluctuations (unwanted bond formations and ligand orientation flipping) restraints were applied for umbrella sampling simulations (Section 2.7.3). Following each simulation free energies for the C-C bond forming step were calculated using the weighted histogram analysis method (WHAM) [Kumar et al., 1992] and plotted in Figure 5.10. Across

Umbrella sampling simulation	eTA MM MD starting trajectory
3*R* Stereochemistry	
1	3*R* MD4 120 psec
2	3*R* MD13 90 psec
3	3*R* MD17 140 psec
3*S* Stereochemistry	
1	3*S* MD4 150 psec
2	3*S* MD16 140 psec
3	3*S* MD18 140 psec
4*	3*S* MD12 130 psec

* Investigated as an individual case

Table 5.3: MM MD simulation trajectories used as starting structures for each of the umbrella sampling simulations.

the three representative umbrella sampling simulations for the 3*R* and 3*S* product forming reactions similar energy maxima were observed. Similar TS energy barriers would suggest poor stereoselectivity in the PS synthesis reaction. This is in agreement with the similiar specificity constants observed for wild-type eTA in Table 4.9.

The proton donor in the mechanism for eTA aldol condensation is proposed to be a conserved active site water molecule [di Salvo et al., 2014]. This is plausible given that simultaneous mutations of proximal histidines (His83 and His126) still yields retro-aldol activity [di Salvo et al., 2014]. We might hope to gain further insight into the eTA mechanism of action through analysis of the QM/MM simulations. As only the ligands shall be treated quantum mechanically we do not expect to see the proton transfer. Rather, we aim to identify locations of suitable catalytic entities when near to the TS structures. Visualisation of each of the three 3*R* and 3*S* product forming simulations are shown in Figure 5.11 and are described in the following sections.

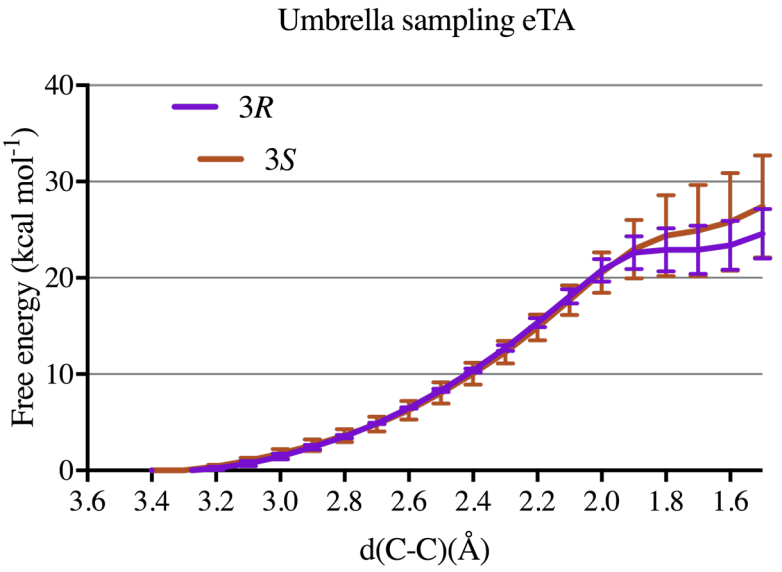


Figure 5.10: WHAM analysis showing the free energies for each of the 3*R* and 3*S* stereochemistries formed along the d(C-C) reaction coordinate at 0.1 Å intervals. Data shown for umbrella sampling from 3*R* MD4,13,17 (purple) and 3*S* MD4,16,18 starting trajectories. Both profiles suggest similar energy maxima. As no proton transfer step is simulated there is no decrease in the energy barrier once the approximate transition state energy is reached.

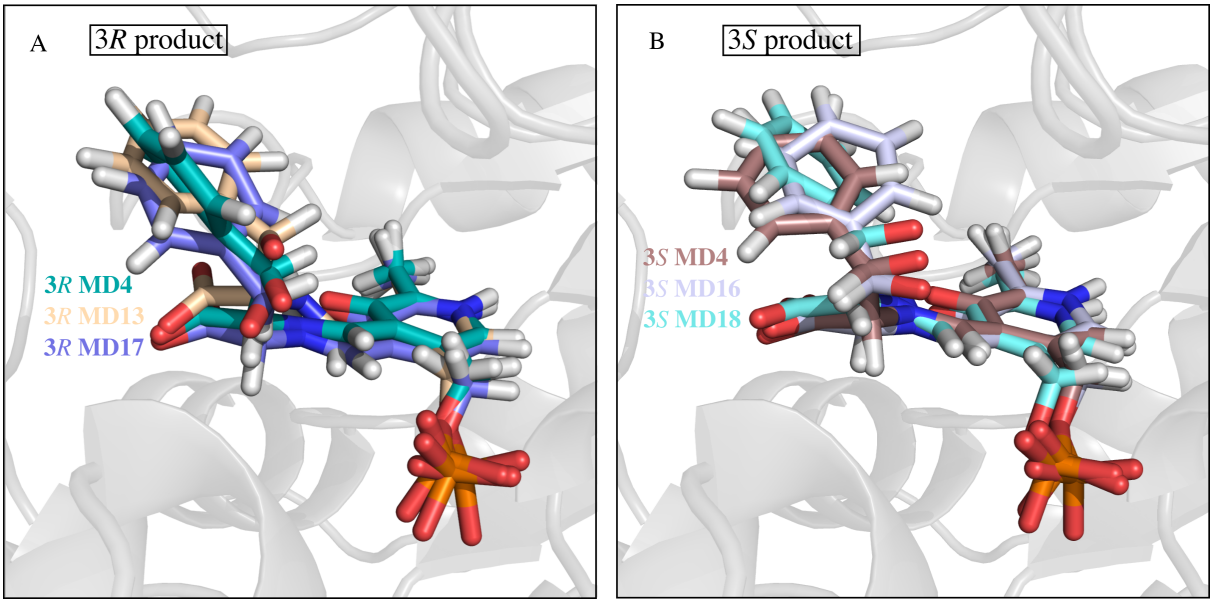


Figure 5.11: Representative umbrella sampling simulations for 3*R* (panel A) and 3*S* (panel B) product forming reactions. Stuctures are shown following C-C bond formation wherby the aldol oxygen is expected to be in position to receive a proton.

5.3.4.1 3*R* product formation

All three umbrella sampling simulations leading to 2*S*3*R* product formation suggest similar binding interactions. Those interactions at the C-C bond forming step are shown in Figure 5.12. No obvious general acid or base residues are in close enough proximity to mediate proton transfer. The most likely candidate for the general acid or base would be His126, however this residue is too far away to make direct contact with aldol oxygen and would require coordination through an active site water molecule. While singly protonated, we might expect His126 to be the general base in the retro-aldol mechanism. This seems unlikely given increases in k_{cat} have been observed when changing this residue to phenylalanine, asparagine [di Salvo et al., 2014][4] and arginine[5](Section 4.5.3). Arg229 coordinates to the aldol oxygen through a well placed water. While not unheard of for arginine to act as a "twisted base" [Schlippe and Hedstrom, 2005, van der Kamp et al., 2008] a pK_a of close to 12 makes this unlikely. Tyr30 coordinates to the aldol oxygen through two water molecules. It therefore seems far more likely that catalytic proton transfer occurs via a Grotth*üss* proton hopping mechanism. This allows for protons to be "shuttled" in and out of the active site through well coordinated water molecules. Interestingly, Y30F and R229N mutations, which would seemingly disrupt this coordination, leads to and increase in 3*S* PS selectivity.

[4]For the retro-aldol cleavage of 2*S*3*R* and 2*S*3*S* threonine
[5]When paired with Y30F R229N mutations for the cleavage of 2*S*3*R* and 2*S*3*S* PS

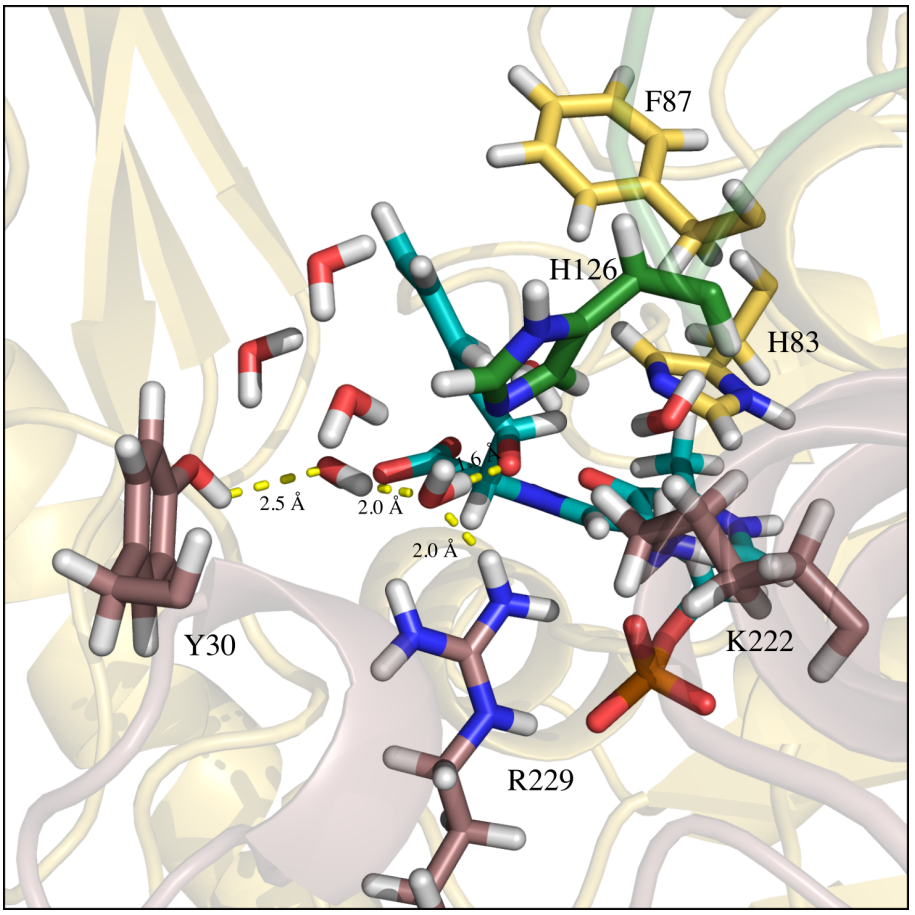


Figure 5.12: A trajectory from umbrella sampling of a 3*R* product forming complex. The 120 psec snapshot from the 3*R* MD4 simulation was used as the starting structure for simulation. The 2*S*3*R* PS-PLP ligand is shown in turquoise where the CA atom of benzaldehyde is within C-C bond forming distance of the CG2 of PXG (1.6 Å). Residues are coloured by their respective chain: Y30, R229N, K229 (brown) chain D, H83 F87 (yellow) chain A and H126 (green) chain B.

5.3.4.2 3*S* product formation

Unlike the umbrella sampling from 3*R* simulations, not all of those leading to a 3*S* stereochemistry displayed similar results. Three representative simulations were observed from starting trajectories in 3*R* MD4, MD16 and MD18 which all had a similar positioning of the aromatic ring upon C-C bond formation (Figure 5.11B). As seen in the previously discussed 3*R* umbrella sampling simulation, Arg229 also coordinates a water molecule in the 3*S* product (Figure 5.13). His126 is in hydrogen bonding distance to coordinate the same water to within 1.8 Å of the aldol oxygen. Lys222 and His83 both orient a separate water to within 1.8 Å of the aldol oxygen. We therefore see two well positioned water molecules in the 3*S* simulation that could mediate proton transfer in the aldol or retro-aldol mechanisms. We therefore propose proton transfer could also occur via the Grotthüss

proton hopping mechanism through the well coordinated water network leading out to bulk solvent.

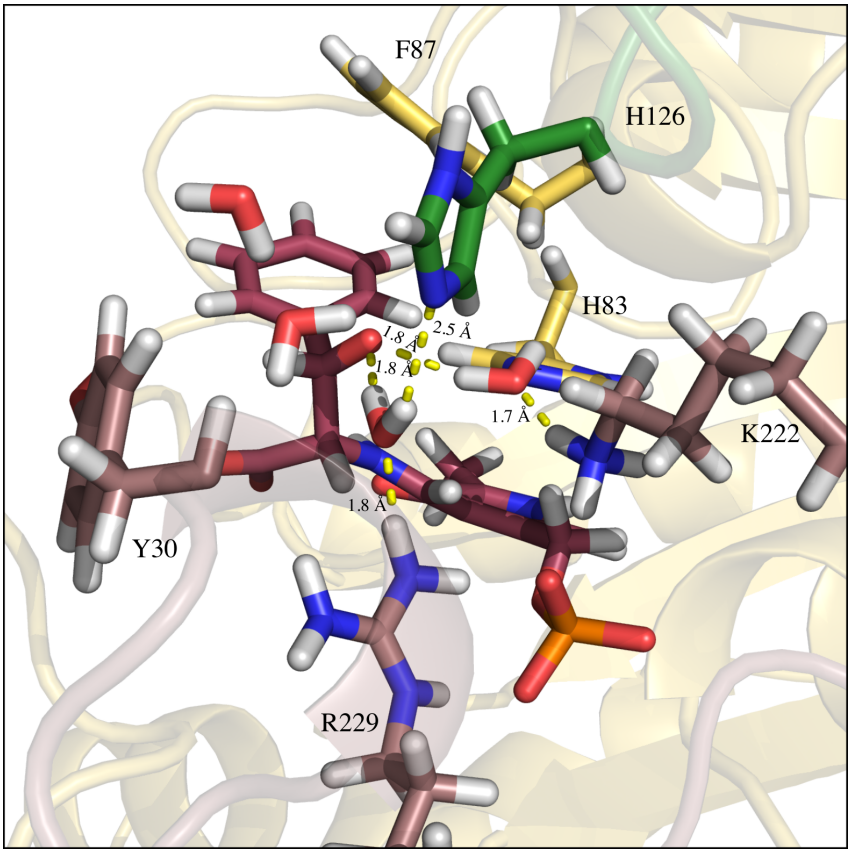


Figure 5.13: A snapshot from umbrella sampling of a 3*S* product forming complex. The 150 psec snapshot from the 3*S* MD4 simulation was used as the starting structure. The 2*S*3*R* PS-PLP ligand is shown in burgundy where the CA atom of benzaldehyde is within C-C bond forming distance of the CG2 of PXG (1.6 Å). Residues are coloured by their respective chain: Y30, R229N, K229 (brown) chain D, H83 F87 (yellow) chain A and H126 (green) chain B.

5.3.5 Suggested mechanisms

In Figures 5.12 and 5.13 we see a representative snapshot of structures close to their product forming transition states. From these simulations we propose the following mechanisms for the stereogenic step in the aldol condensation reaction (Figure 5.14). In the 3*R* product forming reaction Arg229 plays a key role in coordinating a water close to the aldol oxygen (1.6 Å). This water further coordinates to a network of other water molecules leading out to bulk solvent. Abstraction of a proton from the water molecule by benzaldehyde enables *si* face attack to form the aldol C-C bond and subsequent 3*R* product. Protons are able to

shuttle in and out of the active site through the coordinated water network via the Grotthüss proton hopping mechanism.

In the 3*S* product forming reaction an additional water molecule is coordinated to the aldol oxygen through His83 and Lys222. Further, His126 also forms a hydrogen bonding interaction with the same water as Arg229. Abstraction of a proton from either of the two coordinated water molecules is plausible with subsequent *re* face attack forming the 3*S* product. By the same Grotthüss proton hopping mechanism as mentioned above reprotonation of the respective water molecule may occur.

Given these two suggested mechanisms, in addition to the binding analysis data (Section 5.3.3) and WHAM analysis data (Section 5.3.4), it is unsurprising that eTA shows no apparent preference for either 2*S*3*R* or 2*S*3*S* PS. We speculate how a preference for the 3*S* diasteromer arose, during engineering studies (Chapter 4), in Section 5.4.

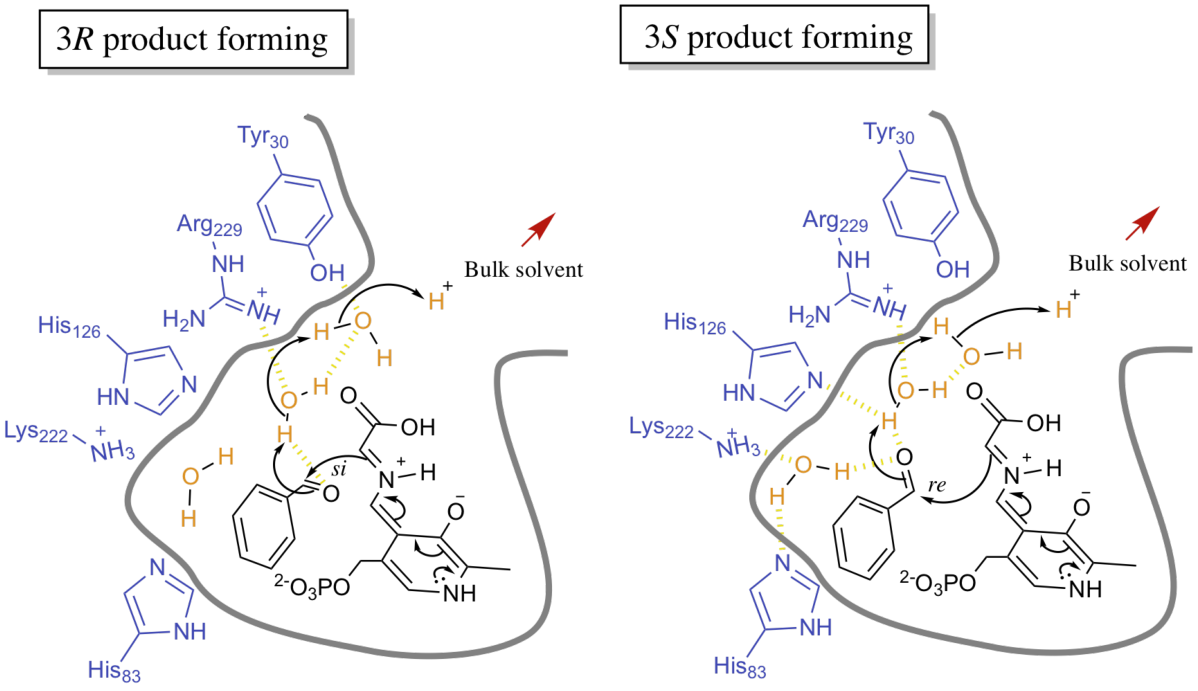


Figure 5.14: A schematic representation of suggested mechanisms for the stereogenic aldol C-C bond forming step. These are given with respect to umbrella sampling trajectories in Figures 5.12 and 5.13. For 3*R* product formation a Arg229 coordinated water molecule allows for proton abstraction and *si* face attack of PXG onto benzaldehyde. A further coordinated water network, involving Tyr30, enables shuttling of protons in and out of the active site via the Grotthüss proton hopping mechanism. In the 3*S* product forming reaction an additional water is coordinated the aldol oxygen of benzaldehyde through His83 and Lys222. This offers an alternative proton for abstraction with subsequent *re* face attack of PXG onto benzaldehyde. Both water molecules offer coordinated networks out to bulk solvent for proton shuttling via the Grotthüss proton hopping mechanism.

5.3.6 An alternative mechanism for 3*S* product formation

Interestingly a different mechanism for aldol protonation is suggested in one of the umbrella sampling simulations. As discussed above, simulations from 3*S* MD4,16 and 18 starting trajectories all displayed similar ring orientations when within C-C bond forming distance. The simulation from the 3*S* MD12 starting trajectory however had an altered ring positioning which directed the carbonyl oxygen of benzaldehyde close enough to the imine of PXG for proton transfer via the mechanism shown in Figure 5.15. WHAM analysis of the umbrella sampling simulation revealed a reduction in the free energy following this transfer, further suggesting this to be a realistic mechanism (Figure 5.16). No other suggestions of such a mechanism have been proposed in the literature, however the pK_a of the imine group in the external aldimine is reported to be close to 8 in a homologous system [Fesko et al., 2018]. This is well suited for such a transfer given all buffers used for kinetic assays are at pH 8. Furthermore a water molecule remains well positioned to reprotonate the imine following transfer to the aldol oxygen.

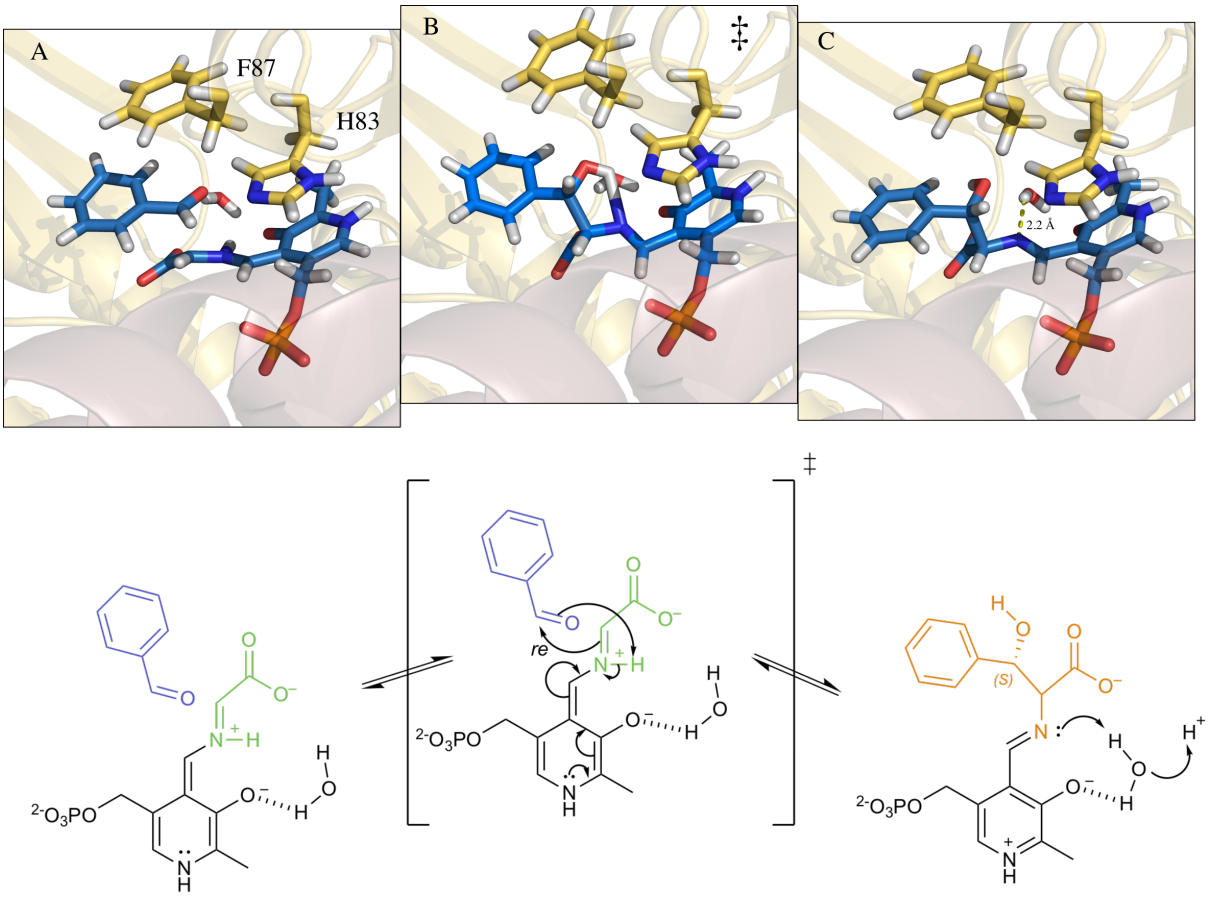


Figure 5.15: The observed proton transfer in umbrella sampling. The simulation was performed with the 3*S* MD12 130 psec starting structure. Panel A and below: Benzaldehyde in a pre-3*S* orientation is angled in a way that brings the carbonyl oxygen into close proximity to the protonated imine of PXG. Panel B and below: The transition state forms whereby the imine proton is abstracted by the benzaldehyde oxygen and PXG attacks onto the CA atom of benzaldehyde forming the C-C bond. Panel C and below: The newly formed hydroxyl proton moves away from the imine and a coordinated water molecule is well placed for reprotonation of the imine.

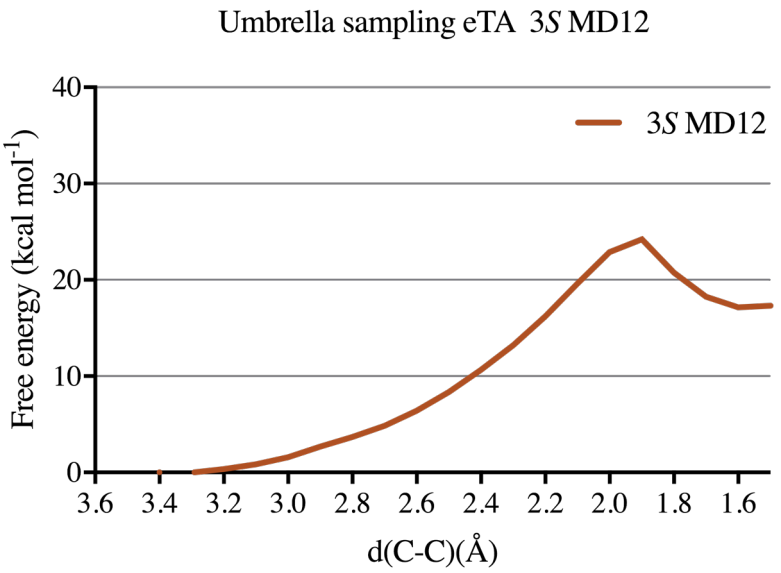


Figure 5.16: WHAM analysis showing the free energy profile for umbrella sampling from a 3*S* MD12 starting trajectory. In the simulation a proton transfer is observed from the imine to the aldol oxygen. Following this a reduction in the free energy is observed defining a clear TS energy barrier at 24 kcal mol^{-1} .

5.4 A rational explanation for improved stereoselectivity in eTA variants

Through CASTing we identified a double Y30F R229N mutant with improved 3*S* selectivity (Chapter 4). The mechanism proposed above for 3*R* product formation (Figure 5.14) requires coordination of a single water molecule by Arg229. Mutation of this residue to an asparagine would make this coordination interaction unlikely; however we might still expect this residue to form a stabilising interaction with the phosphate group of PXG. Without this water molecule coordination, or other obvious route for proton abstraction, it would be unsurprising to see a reduced preference for the 3*R* product. Further, the Y30F mutation would also be expected to disrupt water coordination due to removal of the aromatic hydroxyl group. In the 3*S* forming reaction both His83 and Lys222 coordinate an additional water molecule to the aldol oxygen of benzaldehyde. This doesn't require Arg229 however still forms a network with additional water molecules leading out to bulk solvent. We would therefore still expect the shuttling of protons via the Grotth*ü*ss proton hopping mechanism. Future work should focus on the structural modelling of these, and additional, mutations from Chapter 4 to investigate this.

5.5 Summary

We were able to obtain the highest resolution structure of a L-TA to date at 1.6 Å. This showed covalently bound PLP to Lys197 of the enzyme. The refined eTA structure was used for MM MD simulations to investigate the binding of benzaldehyde in either a pre-3*R* or a pre-3*S* orientation. We further looked at the likelihood of a reactive Michaelis complex forming for each orientation. As might be expected for the poorly diastereoselective enzyme, no preferential binding of a pre-3*R* or -3*S* reactive complex is observed. QM/MM umbrella sampling simulations were performed in an attempt to identify the possible mechanism of proton transfer in the C-C bond forming stereogenic step of the reaction. Representative 3*R* and 3*S* simulations identified well coordinated water networks between bulk solvent and the aldol oxygen suggesting a proton shuttling mechanism. An interesting alternative proton transfer is observed when forming the 3*S* product whereby the aldol oxygen became protonated through the transfer of the imine N_{SB} proton. While three representative umbrella sampling simulations were obtained for the 3*S* product, it is not unusual to observe differences in the transition state structures when different starting trajectories are used [Claeyssens et al., 2011, Lodola et al., 2007]. Thus whilst difficult to prove the feasibility of this mechanism it should not be discounted and makes for an interesting case especially given that all engineering efforts lead to improvements for the 3*S* stereochemistry.

Chapter 6

Summary, Future Work and Perspectives

6.1 Summary

This thesis has demonstrated how a threonine aldolase from *Escherichia coli* (eTA) can be engineered for improved diastereoselectivity with phenylserine (PS) by directed evolution and rational engineering. This was achieved through screening over 3200 transformants with a developed spectrophotmetric screen. High-quality libraries were produced with a focused mutagenesis strategy that allowed for synergistic interactions between residues to be identified. Further, rational combinations of stereoselective variants were made. This identified a quadruple mutant with a 500-fold preference for 2*S*3*S* PS. In this thesis we also present the highest resolution structure of any L-threonine aldolase at 1.6 Å. We used this structure for molecular mechanics (MM) and quantum mechanics/molecular mechanics (QM/MM) simulations to better understand the poor diastereoselective mechanism of the wild-type enzyme. We also present an alternate mechanism in the stereogenic step of the aldol condensation for 2*S*3*S* PS.

In the introduction of this thesis we describe the emergence of biocatalysis, its industrial applications and benefits over chemocatalysis. Further we discuss specific examples of enzymes used to produce sterochemically pure pharmaceutical compounds. One example includes the production of an anti-asthmatic compound with an engineered aminoalcohol dehydrogenase. We then present an in depth discussion of a engineered biocatalyst removing the need for a complex, chemical manufacturing process. This enzyme, a transaminase for the synthesis of sitagliptin, benchmarks modern day efforts in producing

viable biocatalytsts over short time frames. Methods of directed evolution are discussed, these include the generation of diverse gene libraries by error-prone PCR, and homology based recombination. Further we detail the focused mutagenesis strategy of combinatorial active site saturation testing (CASTing) and its success against other directed evolution methods. *De novo* production of artificial enzymes by *in silico* methods are discussed with particular attention paid to an artificial aldolase. In combination with directed evolution methods, this enzyme was engineered to have near natural levels of catalytic efficiency. In addition we discuss the natural aldolases, particularly threonine aldolases (TA) which catalyse the reversible cleavage of threonine. The synthetic uses of these enzymes are presented with their ability to synthesise β-hydroxy-α-amino acids. These compounds contain two chiral centres with four possible stereoisomers and are apparent in antibiotics and as active pharmaceutical ingredients.

In Chapter 3 we clone the gene for a *low-specificity* L-theonine aldolase from *Escherichia coli* into an expression vector. We subsequently transformed this into an expression stain of *E. coli* to obtain large amounts of overexpressed enzyme. This is used to kinetically characterise eTA with 2*S*3*R* and 2*S*3*S* PS revealing the enzyme to be non-selective for either diasteromer. We further developed an assay for the retro-aldol cleavage of PS showing any absorbance between 230-300 nm can be used to follow this activity. This is applied in the development of a high-throughput screen capable of measuring the activity of eTA in crude lysate.

Chapter 4 of the thesis includes directed evolution of eTA. High-quality, focused libraries are generated and screened by the CASTing approach. The developed screen from Chapter 3 is used to identify diastereoselective variants which are further characterised with Michaelis-Menten kinetics. Rational combinations of the diasteroseletive variants yields a quadruple mutant (Y30F R229N H83S F87Y) that has a 500-fold selectivity preference (SP) for 2*S*3*S* PS. This enzyme however shows poor catalytic efficiency due to low activity (k_{cat}) with the selective substrate. Removal of the H83S mutation yields a variant, close to 20-fold selective for 2*S*3*S* PS with a good level of activity. This suggests H83 is important in mediating both activity and diastereoselectivity. By iterative CASTing (ISM), a 10-fold selective variant for for 2*S*3*S* PS is obtained.

In Chapter 5, X-ray crystallography is used to obtain a 1.6 Å structure of eTA. This showed the covalently bound PLP cofactor to Lys197 of the enzyme. The refined eTA structure was used for MM MD simulations to investigate the binding of benzaldehyde in either a pre-3*R* or a pre-3*S* orientation. We further looked at the likelihood of a reactive Michaelis complex forming for each orientation. As might be expected for the poorly diastereoselective enzyme, no preferential binding of a pre-3*R* or -3*S* reactive complex is observed. QM/MM simulations were performed to identify the possible mechanism of proton transfer in the C-C bond forming, stereogenic step of the reaction. Representative 3*R* and 3*S* simulations identified well coordinated water networks between bulk solvent and the aldol oxygen. This suggests proton abstraction may occur via a Grotth*ü*ss hopping mechanism. In this chapter we also present an alternative mechanism for proton transfer when forming the 3*S* product where the aldol oxygen becomes protonated by abstraction of the Schiff base iminium proton. We also speculate how the mutants in Chapter 4 achieve their diastereoselectivity. This involves the disruption of the Arg229 and Tyr30 coordinated water molecules which appear unnecessary for the 3*S* product forming reaction.

6.2 Future work and perspectives

Research continues on the project following submission of this thesis. Efforts will focus on the synthesis of PS, using the identified stereoselective variants. Early efforts have shown proton NMR to be an effective method in identifying peaks corresponding to the formation of 2*S*3*R* and 2*S*3*S* PS. This method will allow for stereoselectivity to be quantified in the aldol condensation reaction. This method could also be used with a range of aldehyde substrates to see if improved stereoselectivity is transferable across different reactions.

The importance of computational modelling is undisputed in modern day enzyme engineering. Future work should focus on modelling the identified enzyme variants *in silico* with MM and QM/MM molecular dynamics. These may also be used to provide insight into additional mutations that could lead to further improvements in stereoselectivity or even a switch in the SP to create 2*S*3*R* PS selective enzymes.

6.3 Concluding remarks

Through this research we have shown that a focused mutagenesis strategy, coupled with an effective screening method, enables improvements in the diastereoselectivity of an enzyme to be achieved. For this we used a plate based high-throughput screening technique. I anticipate this method of screening will become obsolete in the near future given the emergence of ultra-high-throughput, Microfluidics based approaches. With reports of screening up to 5 million variants a day [Ma et al., 2018], the "numbers problem" described in Chapter 1 will likely become more a numbers "hump" as this technology becomes more widely available. In addition, as time progresses so will the advances in computation. We should expect fully robust algorithms that can deliver *de novo* enzymes straight "out the box" and ready for use as effective biocatalysts.

It is fair to say we are at an exciting point in structural and molecular biology. The future is certainly bright for enzymes.

Appendix A

Additional Figures and Tables

Codon Optimised ItaE gene sequence

GAATTCATGGAACATCACCATCACCACCACATTGACCTGCGTTCAGACACCGTTACCCGTCCGTCC
CGTGCTATGCTGGAAGCTATGATGGCTGCTCCGGTTGGCGATGACGTCTATGGTGATGACCCGACC
GTGAACGCCCTGCAGGATTACGCGGCCGAACTGAGTGGTAAAGAAGCAGCTATTTTTCTGCCGACC
GGCACGCAAGCTAATCTGGTGGCGCTGCTGTCCCATTGCGAACGTGGTGAAGAATATATCGTCGGT
CAGGCGGCCCACAATTACCTGTTCGAAGCGGGCGGTGCAGCTGTGCTGGGTAGCATTCAGCCGCAA
CCGATTGATGCCGCGGCAGATGGTACCCTGCCGCTGGATAAAGTTGCTATGAAAATTAAACCGGAT
GACATCCATTTTGCACGTACCAAACTGCTGTCTCTGGAAAACACGCACAATGGCAAAGTCCTGCCG
CGCGAATATCTGAAAGAAGCGTGGGAATTTACCCGTGAACGCAATCTGGCGCTGCATGTTGATGGT
GCACGCATTTTCAATGCCGTGGTTGCATATGGCTGTGAACTGAAAGAAATTACCCAGTACTGCGAC
AGCTTCACGATCTGTCTGTCTAAAGGTCTGGGCACCCCGGTGGGTTCACTGCTGGTTGGCAACCGTG
ATTATATTAAACGCGCCATCCGTTGGCGCAAAATGACGGGCGGTGGCATGCGTCAATCGGGTATTCT
GGCTGCGGCCGGCATCTACGCTCTGAAAAACAATGTGGCGCGTCTGCAGGAAGATCACGACAATGC
AGCTTGGATGGCGGAACAGCTGCGTGAAGCTGGTGCGGATGTTATGCGCCAAGACACCAACATGCT
GTTTGTCCGTGTGGGTGAAGAAAATGCGGCCGCACTGGGCGAATACATGAAAGCCCGTAACGTGCT
GATTAATGCATCCCCGATCGTTCGTCTGGTGACGCATCTGGACGTGAGCCGTGAACAACTGGCGGAA
GTCGCTGCCCACTGGCGTGCGTTTCTGGCGCGTTAATAAAAGCTT

Figure A.1: Codon optimised ItaE gene sequence. The gene was designed to incorporate an N-terminal His-tag into the enzyme and is 1038 bp in length.

Name	Sequence
ItaE forward	5' - ACAGGAAACAGAATTC ATGGAACATCACCATCACCACCAC -3'
ItaE reverse	5' - AGAAGCTTGGCTGCAG TTATTAACGCGCCAGAAACGCACG-3'
pKK223-3 forward	5' - CTGCAGCCAAGCTTCTGTTTTG -3'
pKK223-3 reverse	5' - GAATTCTGTTTCCTGTGTAAATTGTTATCC -3'

Table A.1: Primer sequences for ligation independent cloning. Primers were designed to contain complementary overlapping regions for the gene and plasmid using the methods described by Li et al. [Li et al., 2011].

Name	Sequence
CASTing	
Group 1 forward	5’ - GTCAGGCGGCCNDTAATTACCTGNDTGAAGCGGGCGG -3’
Group 1 reverse	5’ - CCGCCCGCTTCAHNCAGGTAATTAHNGGCCGCCTGAC -3’
Group 2 forward	5’ - CCGGATGACATCNDTNDTGCACGTACCAAACTGC -3’
Group 2 reverse	5’ - GCAGTTTGGTACGTGCAHNAHNGATGTCATCCGG -3’
Group 3 forward	5’ - GCACCGCCCGCTTCGAAAHNGTAATTGTGGGCAHNCTGACC -3’
Group 3 reverse	5’ - GGTCAGNDTGCCCACAATTACNDTTTCGAAGCGGGCGGTGC -3’
Group 4 forward 1	5’ - GTTGGCGATGACGTCNDTGGTGATGACCCGACC -3’
Group 4 forward 2	5’ - GACGGGCGGTGGCATGNDTCAATCGGGTATTCTGG -3’
Rational combinants 1	
H83S F87Y forward	5’-GTCAGGCGGGCCTCTAATTACCTGTATGAAGCGGGCGG-3’
H83S F87Y reverse	5’-CCGCCCGCTTCATACAGGTAATTAGAGGCCCGCCTGAC-3’
H83S forward	5’-CTTCGAACAGGTAATTGCTGGCCGCCTGACCGACG 3’
H83S reverse	5’-CGTCGGTCAGGCGGCCAGCAATTACCTGTTCGAAG-3’
F87Y forward	5’-TGCACCGCCCGCTTCATACAGGTAATTGTGGGC-3’
F87Y reverse	5’-GCCCACAATTACCTGTATGAAGCGGGCGGTGCA-3’
Rational combinants 2	
H126R forward	5’-CAGTTTGGTACGTGCAAAACGGATGTCATCCGGTTTAAT- 3’
H126R reverse	5’-ATTAAACCGGATGACATCCGTTTTGCACGTACCAAACTG-3’
F127H forward	5’-AGCAGTTTGGTACGTGCATGATGGATGTCATCCGGTTTAATTTTC-3’
F127H reverse	5’-GAAAATTAAACCGGATGACATCCATCATGCACGTACCAAACTGCT-3’
H126R F127H forward	5’-CCGGATGACATCCGTCATGCACGTACCAAACTGC-3’
H126R F127H reverse	5’-GCAGTTTGGTACGTGCATGACGGATGTCATCCGG-3’

Table A.2: Primers for eTA CASTing groups and those used to make rational combinants.

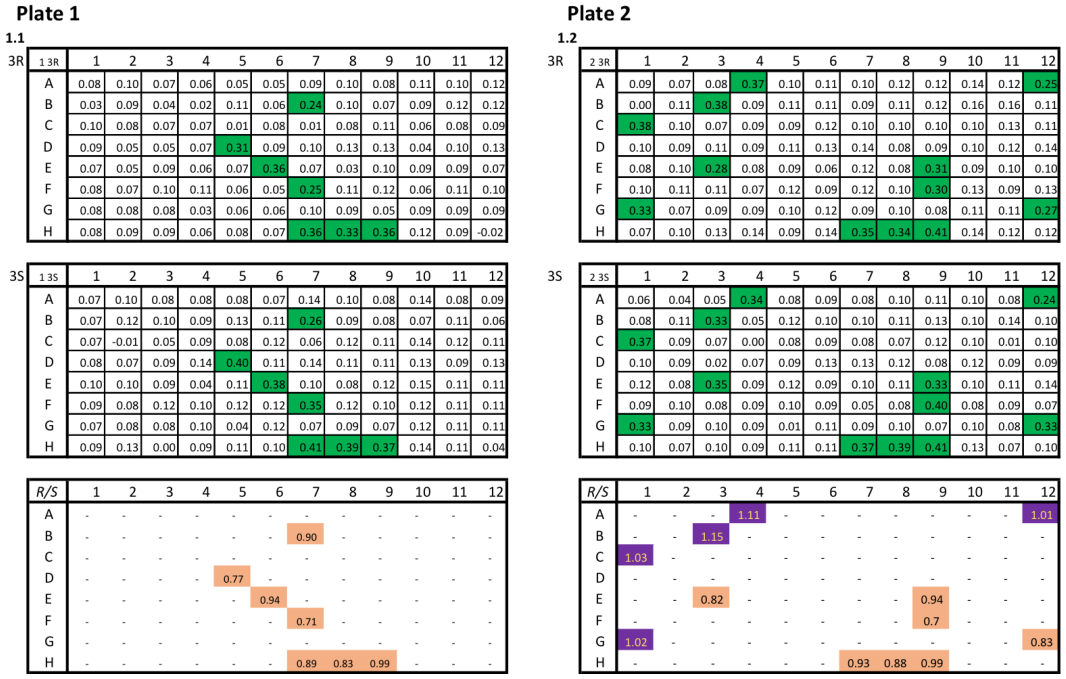


Figure A.2: Screening plates for CASTing group 1 ($H83_{NDT}$ $F87_{NDT}$) library plates 1 (left) and 2 (right). For each library plate containing eTA variants, two screening plates are set up identically to measure retro-aldol activity of either 2*S*3*R* (top) or 2*S*3*S* (middle) PS. Positive controls are shown in wells H7-9 and negative controls in wells H10-12. The activity of each well is measured from the change in Abs_{260}. Those wells with an activity of 0.2 ΔAbs_{260} or greater are highlighted in green. Identifier analysis (bottom) shows the 2*S*3*R* / 2*S*3*S* activity ratio for wells highlighted in green above. Those in purple will have greater levels of activity for the 2*S*3*R* PS retro-aldol cleavage and those in brown for 2*S*3*S*.

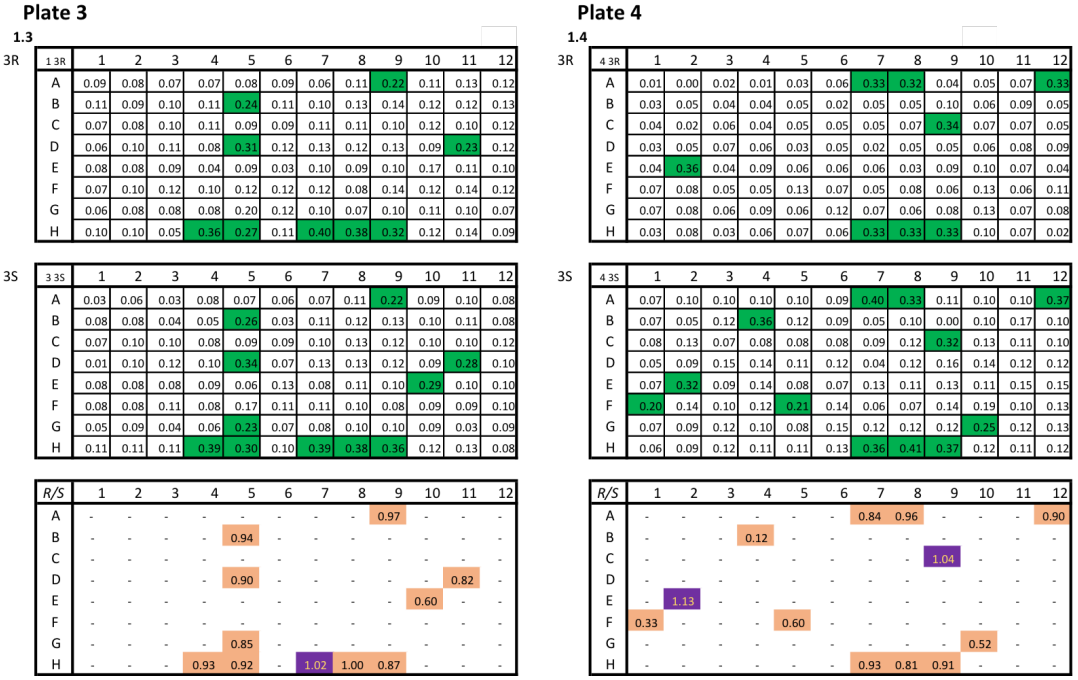


Figure A.3: As Figure A.2 for library plates 3 (left) and 4 (right).

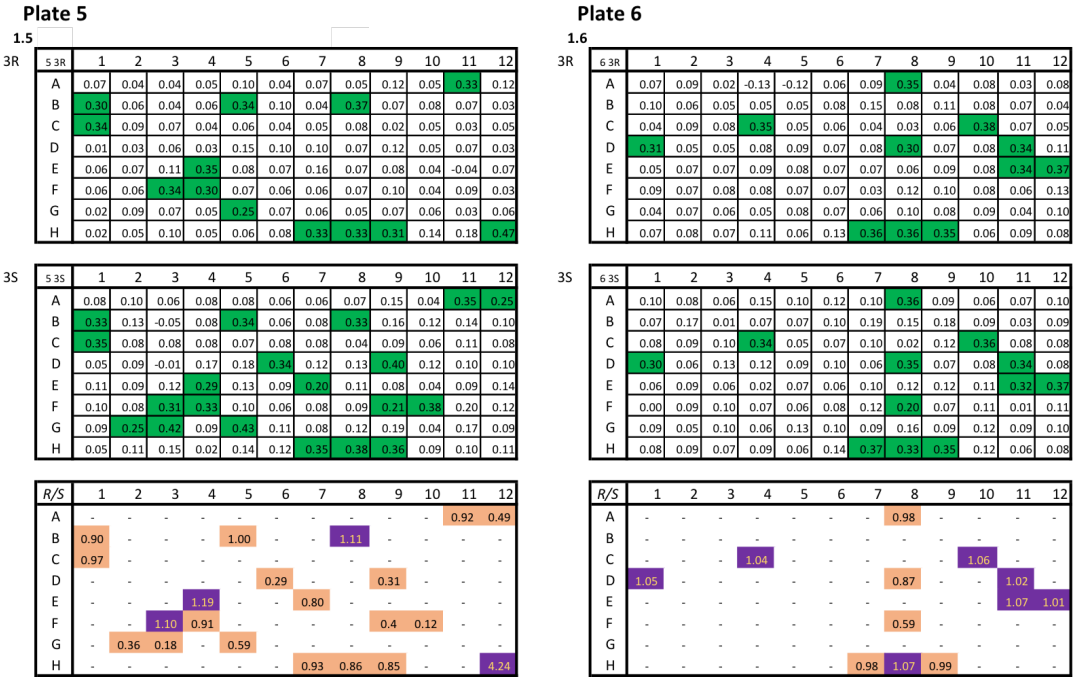


Figure A.4: As Figure A.2 for library plates 5 (left) and 6 (right).

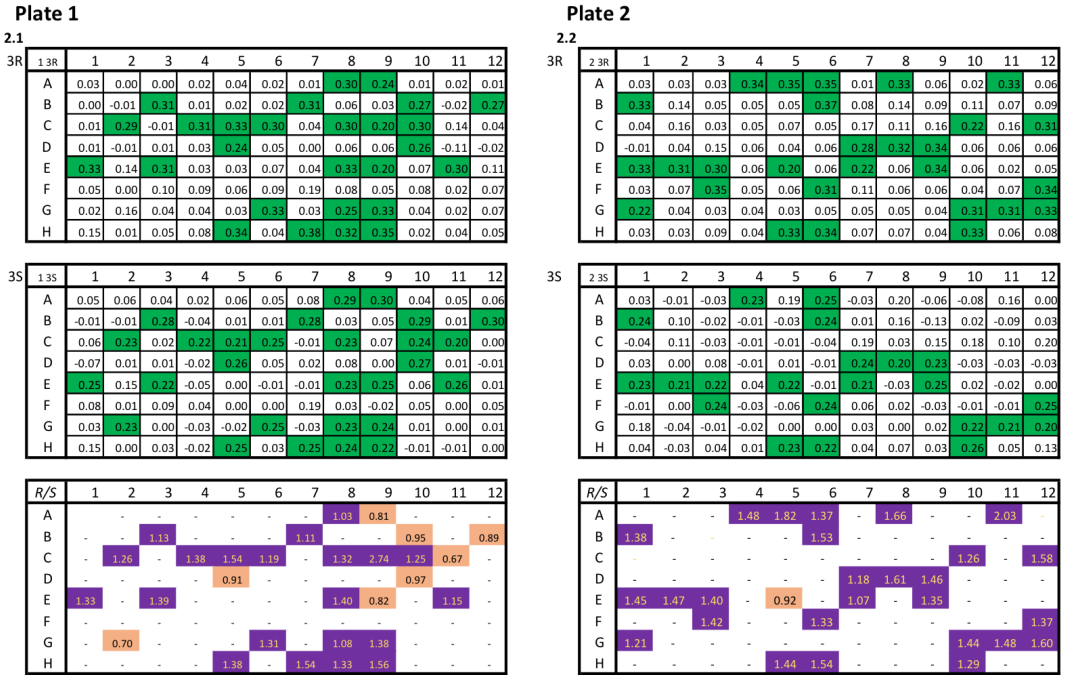


Figure A.5: Screening plates for CASTing group 2 ($H126_{NDT}$ $F127_{NDT}$) library plates 1 (left) and 2 (right). For each library plate containing eTA variants, two screening plates are set up identically to measure retro-aldol activity of either 2*S*3*R* (top) or 2*S*3*S* (middle) PS. Positive controls are shown in wells H7-9 and negative controls in wells H10-12. The activity of each well is measured from the change in Abs_{260}. Those wells with an activity of 0.2 ΔAbs_{260} or greater are highlighted in green. Identifier analysis (bottom) shows the 2*S*3*R* / 2*S*3*S* activity ratio for wells highlighted in green above. Those in purple will have greater levels of activity for the 2*S*3*R* PS retro-aldol cleavage and those in brown for 2*S*3*S*.

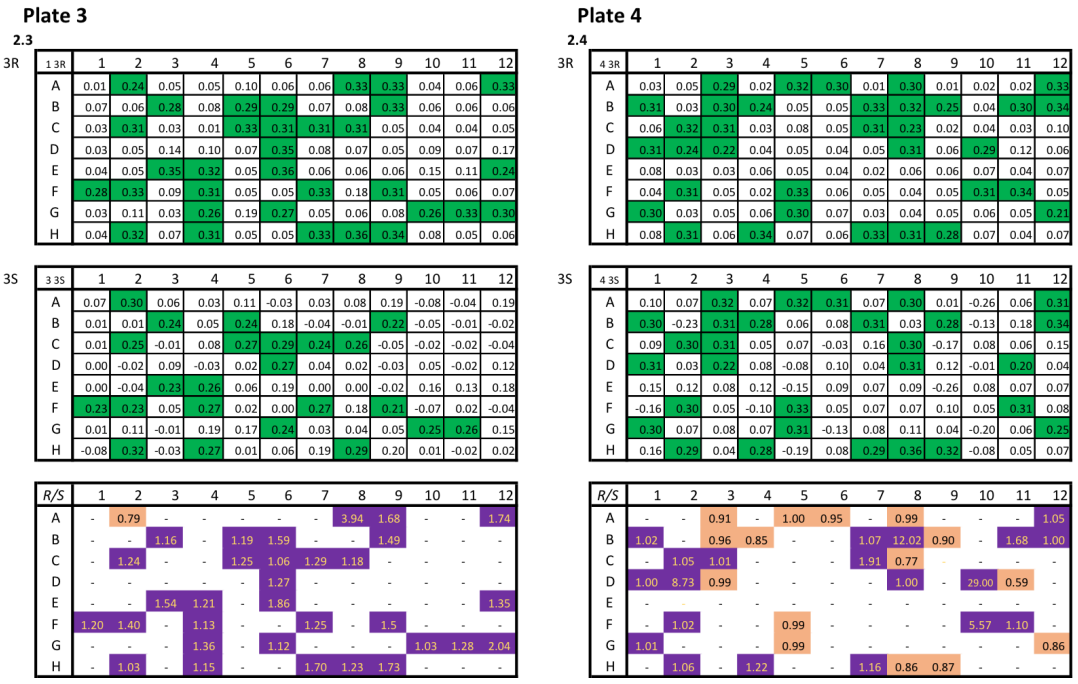


Figure A.6: As Figure A.5 for library plates 3 (left) and 4 (right).

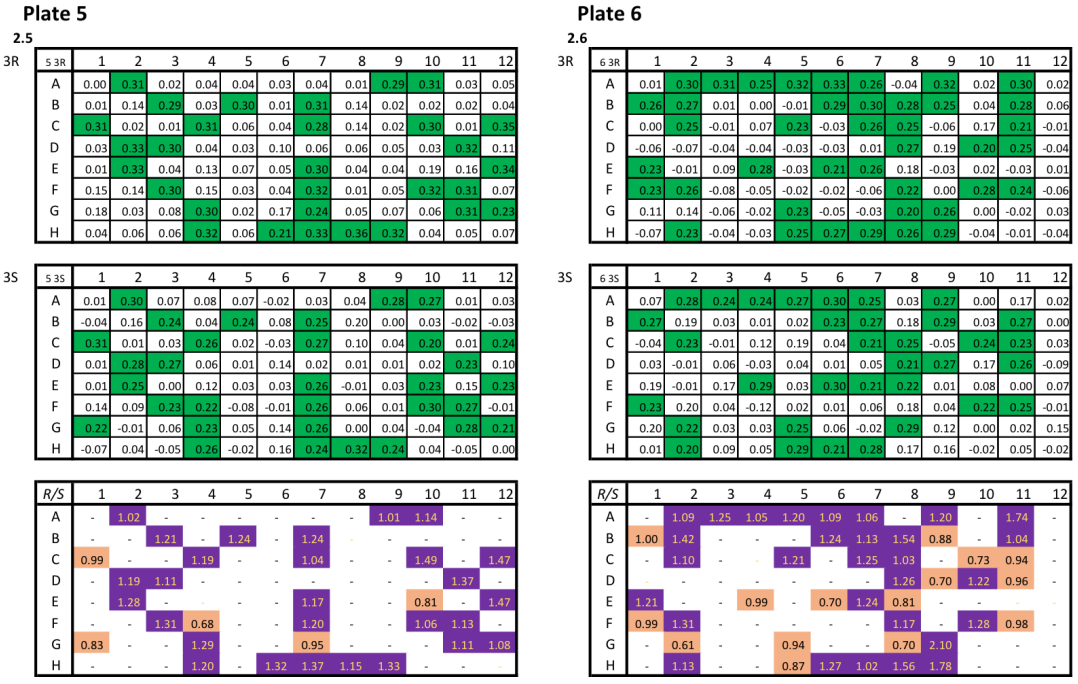


Figure A.7: As Figure A.5 for library plates 5 (left) and 6 (right).

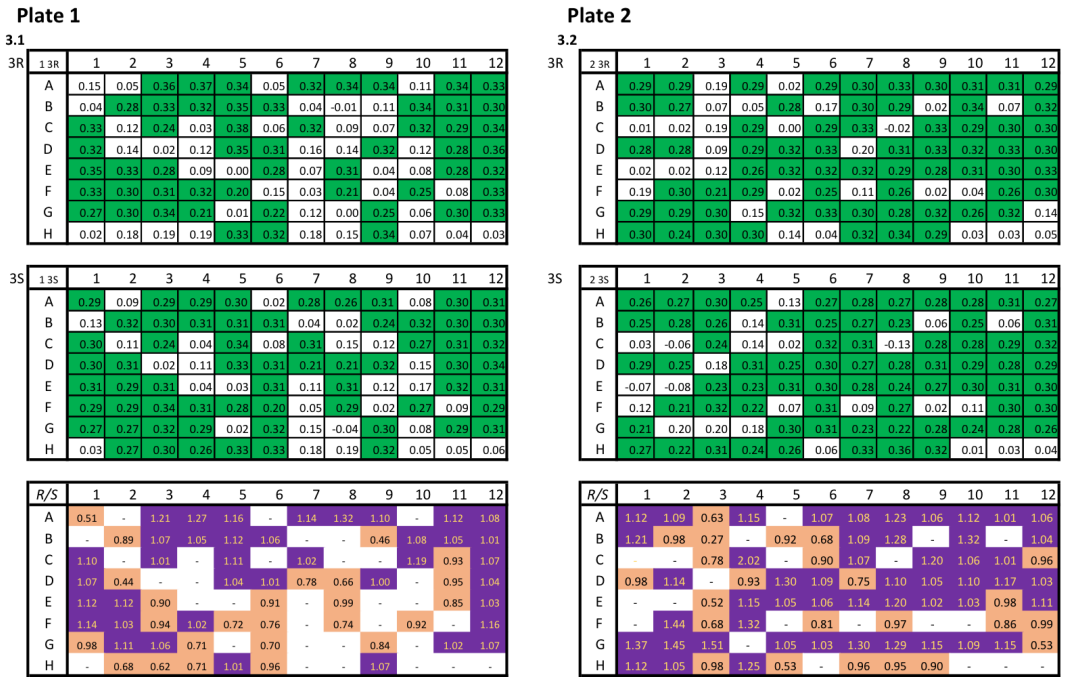


Plate 1

3.1

3R	1 3R	1	2	3	4	5	6	7	8	9	10	11	12
	A	0.15	0.05	0.36	0.37	0.34	0.05	0.32	0.34	0.34	0.11	0.34	0.33
	B	0.04	0.28	0.33	0.32	0.35	0.33	0.04	-0.01	0.11	0.34	0.31	0.30
	C	0.33	0.12	0.24	0.03	0.38	0.06	0.32	0.09	0.07	0.32	0.29	0.34
	D	0.32	0.14	0.02	0.12	0.35	0.31	0.16	0.14	0.32	0.12	0.28	0.36
	E	0.35	0.33	0.28	0.09	0.00	0.28	0.07	0.31	0.04	0.08	0.28	0.32
	F	0.33	0.30	0.31	0.32	0.20	0.15	0.03	0.21	0.04	0.25	0.08	0.33
	G	0.27	0.30	0.34	0.21	0.01	0.22	0.12	0.00	0.25	0.06	0.30	0.33
	H	0.02	0.18	0.19	0.19	0.33	0.32	0.18	0.15	0.34	0.07	0.04	0.03

3S	1 3S	1	2	3	4	5	6	7	8	9	10	11	12
	A	0.29	0.09	0.29	0.29	0.30	0.02	0.28	0.26	0.31	0.08	0.30	0.31
	B	0.13	0.32	0.30	0.31	0.31	0.31	0.04	0.02	0.24	0.32	0.30	0.30
	C	0.30	0.11	0.24	0.04	0.34	0.08	0.31	0.15	0.12	0.27	0.31	0.32
	D	0.30	0.31	0.02	0.11	0.33	0.31	0.21	0.21	0.32	0.15	0.30	0.34
	E	0.31	0.29	0.31	0.04	0.03	0.31	0.11	0.31	0.12	0.17	0.32	0.31
	F	0.29	0.29	0.34	0.31	0.28	0.20	0.05	0.29	0.02	0.27	0.09	0.29
	G	0.27	0.27	0.32	0.29	0.02	0.32	0.15	-0.04	0.30	0.08	0.29	0.31
	H	0.03	0.27	0.30	0.26	0.33	0.33	0.18	0.19	0.32	0.05	0.05	0.06

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.51	-	1.21	1.27	1.16	-	1.14	1.32	1.10	-	1.12	1.08
B	-	0.89	1.07	1.05	1.12	1.06	-	-	0.46	1.08	1.05	1.01
C	1.10	-	1.01	-	1.11	-	1.02	-	-	1.19	0.93	1.07
D	1.07	0.44	-	-	1.04	1.01	0.78	0.66	1.00	-	0.95	1.04
E	1.12	1.12	0.90	-	-	0.91	-	0.99	-	-	0.85	1.03
F	1.14	1.03	0.94	1.02	0.72	0.76	-	0.74	-	0.92	-	1.16
G	0.98	1.11	1.06	0.71	-	0.70	-	-	0.84	-	1.02	1.07
H	-	0.68	0.62	0.71	1.01	0.96	-	-	1.07	-	-	-

Plate 2

3.2

3R	2 3R	1	2	3	4	5	6	7	8	9	10	11	12
	A	0.29	0.29	0.19	0.29	0.02	0.29	0.30	0.33	0.30	0.31	0.31	0.29
	B	0.30	0.27	0.07	0.05	0.28	0.17	0.30	0.29	0.02	0.34	0.07	0.32
	C	0.01	0.02	0.19	0.29	0.00	0.29	0.33	-0.02	0.33	0.29	0.30	0.30
	D	0.28	0.28	0.09	0.29	0.32	0.33	0.20	0.31	0.33	0.32	0.33	0.30
	E	0.02	0.02	0.12	0.26	0.32	0.32	0.32	0.29	0.28	0.31	0.30	0.33
	F	0.19	0.30	0.21	0.29	0.02	0.25	0.11	0.26	0.02	0.04	0.26	0.30
	G	0.29	0.29	0.30	0.15	0.32	0.33	0.30	0.28	0.32	0.26	0.32	0.14
	H	0.30	0.24	0.30	0.30	0.14	0.04	0.32	0.34	0.29	0.03	0.03	0.05

3S	2 3S	1	2	3	4	5	6	7	8	9	10	11	12
	A	0.26	0.27	0.30	0.25	0.13	0.27	0.28	0.27	0.28	0.28	0.31	0.27
	B	0.25	0.28	0.26	0.14	0.31	0.25	0.27	0.23	0.06	0.25	0.06	0.31
	C	0.03	-0.06	0.24	0.14	0.02	0.32	0.31	-0.13	0.28	0.28	0.29	0.32
	D	0.29	0.25	0.18	0.31	0.25	0.30	0.27	0.28	0.31	0.29	0.28	0.29
	E	-0.07	-0.08	0.23	0.23	0.31	0.30	0.28	0.24	0.27	0.30	0.31	0.30
	F	0.12	0.21	0.32	0.22	0.07	0.31	0.09	0.27	0.02	0.11	0.30	0.30
	G	0.21	0.20	0.20	0.18	0.30	0.31	0.23	0.22	0.28	0.24	0.28	0.26
	H	0.27	0.22	0.31	0.24	0.26	0.06	0.33	0.36	0.32	0.01	0.03	0.04

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	1.12	1.09	0.63	1.15	-	1.07	1.08	1.23	1.06	1.12	1.01	1.06
B	1.21	0.98	0.27	-	0.92	0.68	1.09	1.28	-	1.32	-	1.04
C	-	-	0.78	2.02	-	0.90	1.07	-	1.20	1.06	1.01	0.96
D	0.98	1.14	-	0.93	1.30	1.09	0.75	1.10	1.05	1.10	1.17	1.03
E	-	-	0.52	1.15	1.05	1.06	1.14	1.20	1.02	1.03	0.98	1.11
F	-	1.44	0.68	1.32	-	0.81	-	0.97	-	-	0.86	0.99
G	1.37	1.45	1.51	-	1.05	1.03	1.30	1.29	1.15	1.09	1.15	0.53
H	1.12	1.05	0.98	1.25	0.53	-	0.96	0.95	0.90	-	-	-

Figure A.8: Screening plates for CASTing group 3 ($A81_{NDT}$ $L86_{NDT}$) library plates 1 (left) and 2 (right). For each library plate containing eTA variants, two screening plates are set up identically to measure retro-aldol activity of either 2*S*3*R* (top) or 2*S*3*S* (middle) PS. Positive controls are shown in wells H7-9 and negative controls in wells H10-12. The activity of each well is measured from the change in Abs_{260}. Those wells with an activity of 0.2 ΔAbs_{260} or greater are highlighted in green. Identifier analysis (bottom) shows the 2*S*3*R* / 2*S*3*S* activity ratio for wells highlighted in green above. Those in purple will have greater levels of activity for the 2*S*3*R* PS retro-aldol cleavage and those in brown for 2*S*3*S*.

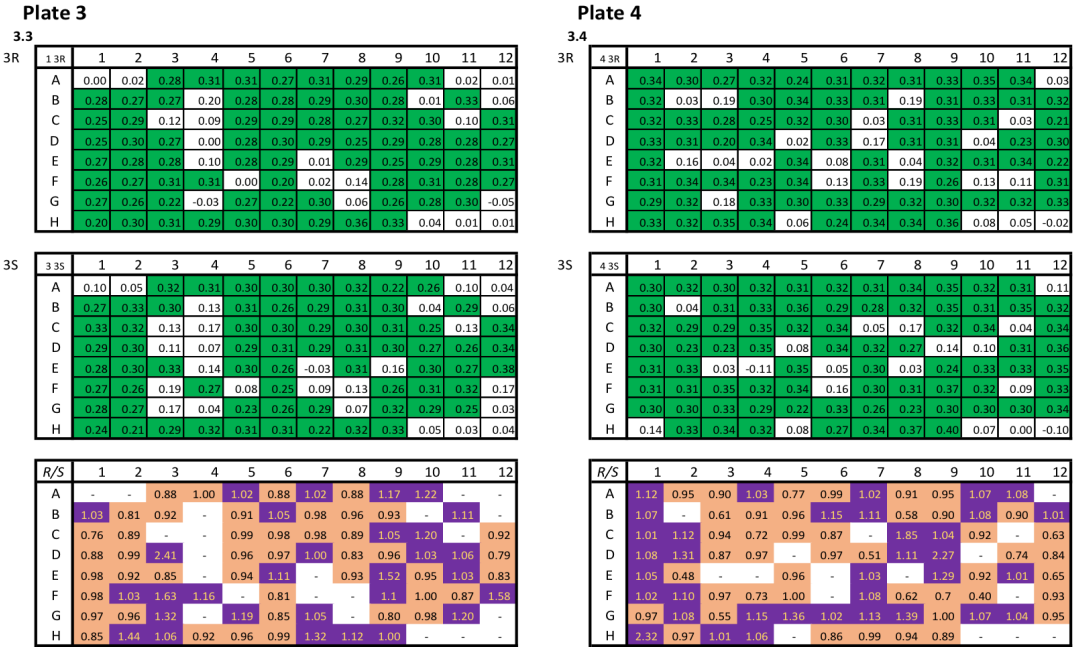


Plate 3

3.3

3R	1 3R	1	2	3	4	5	6	7	8	9	10	11	12
	A	0.00	0.02	0.28	0.31	0.31	0.27	0.31	0.29	0.26	0.31	0.02	0.01
	B	0.28	0.27	0.27	0.20	0.28	0.28	0.29	0.30	0.28	0.01	0.33	0.06
	C	0.25	0.29	0.12	0.09	0.29	0.29	0.28	0.27	0.32	0.30	0.10	0.31
	D	0.25	0.30	0.27	0.00	0.28	0.30	0.29	0.25	0.29	0.28	0.28	0.27
	E	0.27	0.28	0.28	0.10	0.28	0.29	0.01	0.29	0.25	0.29	0.28	0.31
	F	0.26	0.27	0.31	0.31	0.00	0.20	0.02	0.14	0.28	0.31	0.28	0.27
	G	0.27	0.26	0.22	-0.03	0.27	0.22	0.30	0.06	0.26	0.28	0.30	-0.05
	H	0.20	0.30	0.31	0.29	0.30	0.30	0.29	0.36	0.33	0.04	0.01	0.01

3S	3 3S	1	2	3	4	5	6	7	8	9	10	11	12
	A	0.10	0.05	0.32	0.31	0.30	0.30	0.30	0.32	0.22	0.26	0.10	0.04
	B	0.27	0.33	0.30	0.13	0.31	0.26	0.29	0.31	0.30	0.04	0.29	0.06
	C	0.33	0.32	0.13	0.17	0.30	0.30	0.29	0.30	0.31	0.25	0.13	0.34
	D	0.29	0.30	0.11	0.07	0.29	0.31	0.29	0.31	0.30	0.27	0.26	0.34
	E	0.28	0.30	0.33	0.14	0.30	0.26	-0.03	0.31	0.16	0.30	0.27	0.38
	F	0.27	0.26	0.19	0.27	0.08	0.25	0.09	0.13	0.26	0.31	0.32	0.17
	G	0.28	0.27	0.17	0.04	0.23	0.26	0.29	0.07	0.32	0.29	0.25	0.03
	H	0.24	0.21	0.29	0.32	0.31	0.31	0.22	0.32	0.33	0.05	0.03	0.04

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	0.88	1.00	1.02	0.88	1.02	0.88	1.17	1.22	-	-
B	1.03	0.81	0.92	-	0.91	1.05	0.98	0.96	0.93	-	1.11	-
C	0.76	0.89	-	-	0.99	0.98	0.98	0.89	1.05	1.20	-	0.92
D	0.88	0.99	2.41	-	0.96	0.97	1.00	0.83	0.96	1.03	1.06	0.79
E	0.98	0.92	0.85	-	0.94	1.11	-	0.93	1.52	0.95	1.03	0.83
F	0.98	1.03	1.63	1.16	-	0.81	-	-	1.1	1.00	0.87	1.58
G	0.97	0.96	1.32	-	1.19	0.85	1.05	-	0.80	0.98	1.20	-
H	0.85	1.44	1.06	0.92	0.96	0.99	1.32	1.12	1.00	-	-	-

Plate 4

3.4

3R	4 3R	1	2	3	4	5	6	7	8	9	10	11	12
	A	0.34	0.30	0.27	0.32	0.24	0.31	0.32	0.31	0.33	0.35	0.34	0.03
	B	0.32	0.03	0.19	0.30	0.34	0.33	0.31	0.19	0.31	0.33	0.31	0.32
	C	0.32	0.33	0.28	0.25	0.32	0.30	0.03	0.31	0.33	0.31	0.03	0.21
	D	0.33	0.31	0.20	0.34	0.02	0.33	0.17	0.31	0.31	0.04	0.23	0.30
	E	0.32	0.16	0.04	0.02	0.34	0.08	0.31	0.04	0.32	0.31	0.34	0.22
	F	0.31	0.34	0.34	0.23	0.34	0.13	0.33	0.19	0.26	0.13	0.11	0.31
	G	0.29	0.32	0.18	0.33	0.30	0.33	0.29	0.32	0.30	0.32	0.32	0.33
	H	0.33	0.32	0.35	0.34	0.06	0.24	0.34	0.34	0.36	0.08	0.05	-0.02

3S	4 3S	1	2	3	4	5	6	7	8	9	10	11	12
	A	0.30	0.32	0.30	0.32	0.31	0.32	0.31	0.34	0.35	0.32	0.31	0.11
	B	0.30	0.04	0.31	0.33	0.36	0.29	0.28	0.32	0.35	0.31	0.35	0.32
	C	0.32	0.29	0.29	0.35	0.32	0.34	0.05	0.17	0.32	0.34	0.04	0.34
	D	0.30	0.23	0.23	0.35	0.08	0.34	0.32	0.27	0.14	0.10	0.31	0.36
	E	0.31	0.33	0.03	-0.11	0.35	0.05	0.30	0.03	0.24	0.33	0.33	0.35
	F	0.31	0.31	0.35	0.32	0.34	0.16	0.30	0.31	0.37	0.32	0.09	0.33
	G	0.30	0.30	0.33	0.29	0.22	0.33	0.26	0.23	0.30	0.30	0.30	0.34
	H	0.14	0.33	0.34	0.32	0.08	0.27	0.34	0.37	0.40	0.07	0.00	-0.10

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	1.12	0.95	0.90	1.03	0.77	0.99	1.02	0.91	0.95	1.07	1.08	-
B	1.07	-	0.61	0.91	0.96	1.15	1.11	0.58	0.90	1.08	0.90	1.01
C	1.01	1.12	0.94	0.72	0.99	0.87	-	1.85	1.04	0.92	-	0.63
D	1.08	1.31	0.87	0.97	-	0.97	0.51	1.11	2.27	-	0.74	0.84
E	1.05	0.48	-	-	0.96	-	1.03	-	1.29	0.92	1.01	0.65
F	1.02	1.10	0.97	0.73	1.00	-	1.08	0.62	0.7	0.40	-	0.93
G	0.97	1.08	0.55	1.15	1.36	1.02	1.13	1.39	1.00	1.07	1.04	0.95
H	2.32	0.97	1.01	1.06	-	0.86	0.99	0.94	0.89	-	-	-

Figure A.9: As Figure A.8 for library plates 3 (left) and 4 (right).

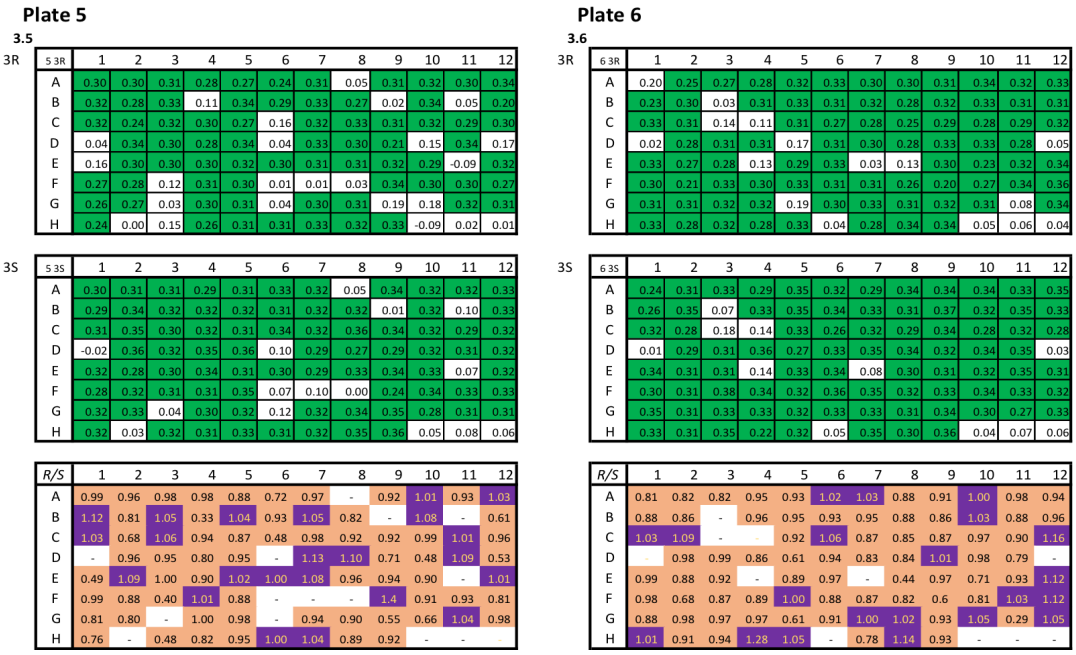


Plate 5

3.5 — 3R

5 3R	1	2	3	4	5	6	7	8	9	10	11	12
A	0.30	0.30	0.31	0.28	0.27	0.24	0.31	0.05	0.31	0.32	0.30	0.34
B	0.32	0.28	0.33	0.11	0.34	0.29	0.33	0.27	0.02	0.34	0.05	0.20
C	0.32	0.24	0.32	0.30	0.27	0.16	0.32	0.33	0.31	0.32	0.29	0.30
D	0.04	0.34	0.30	0.28	0.34	0.04	0.33	0.30	0.21	0.15	0.34	0.17
E	0.16	0.30	0.30	0.30	0.32	0.30	0.31	0.31	0.32	0.29	-0.09	0.32
F	0.27	0.28	0.12	0.31	0.30	0.01	0.01	0.03	0.34	0.30	0.30	0.27
G	0.26	0.27	0.03	0.30	0.31	0.04	0.30	0.31	0.19	0.18	0.32	0.31
H	0.24	0.00	0.15	0.26	0.31	0.31	0.33	0.32	0.33	-0.09	0.02	0.01

3S

5 3S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.30	0.31	0.31	0.29	0.31	0.33	0.32	0.05	0.34	0.32	0.32	0.33
B	0.29	0.34	0.32	0.32	0.32	0.31	0.32	0.32	0.01	0.32	0.10	0.33
C	0.31	0.35	0.30	0.32	0.31	0.34	0.32	0.36	0.34	0.32	0.29	0.32
D	-0.02	0.36	0.32	0.35	0.36	0.10	0.29	0.27	0.29	0.32	0.31	0.32
E	0.32	0.28	0.30	0.34	0.31	0.30	0.29	0.33	0.34	0.33	0.07	0.32
F	0.28	0.32	0.31	0.31	0.35	0.07	0.10	0.00	0.24	0.34	0.33	0.33
G	0.32	0.33	0.04	0.30	0.32	0.12	0.32	0.34	0.35	0.28	0.31	0.31
H	0.32	0.03	0.32	0.31	0.33	0.31	0.32	0.35	0.36	0.05	0.08	0.06

R/S

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.99	0.96	0.98	0.98	0.88	0.72	0.97	-	0.92	1.01	0.93	1.03
B	1.12	0.81	1.05	0.33	1.04	0.93	1.05	0.82	-	1.08	-	0.61
C	1.03	0.68	1.06	0.94	0.87	0.48	0.98	0.92	0.92	0.99	1.01	0.96
D	-	0.96	0.95	0.80	0.95	-	1.13	1.10	0.71	0.48	1.09	0.53
E	0.49	1.09	1.00	0.90	1.02	1.00	1.08	0.96	0.94	0.90	-	1.01
F	0.99	0.88	0.40	1.01	0.88	-	-	-	1.4	0.91	0.93	0.81
G	0.81	0.80	-	1.00	0.98	-	0.94	0.90	0.55	0.66	1.04	0.98
H	0.76	-	0.48	0.82	0.95	1.00	1.04	0.89	0.92	-	-	-

Plate 6

3.6 — 3R

6 3R	1	2	3	4	5	6	7	8	9	10	11	12
A	0.20	0.25	0.27	0.28	0.32	0.33	0.30	0.30	0.31	0.34	0.32	0.33
B	0.23	0.30	0.03	0.31	0.33	0.31	0.32	0.28	0.32	0.33	0.31	0.31
C	0.33	0.31	0.14	0.11	0.31	0.27	0.28	0.25	0.29	0.28	0.29	0.32
D	0.02	0.28	0.31	0.31	0.17	0.31	0.30	0.28	0.33	0.33	0.28	0.05
E	0.33	0.27	0.28	0.13	0.29	0.33	0.03	0.13	0.30	0.23	0.32	0.34
F	0.30	0.21	0.33	0.30	0.33	0.31	0.31	0.26	0.20	0.27	0.34	0.36
G	0.31	0.31	0.32	0.32	0.19	0.30	0.33	0.31	0.32	0.31	0.08	0.34
H	0.33	0.28	0.32	0.28	0.33	0.04	0.28	0.34	0.34	0.05	0.06	0.04

3S

6 3S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.24	0.31	0.33	0.29	0.35	0.32	0.29	0.34	0.34	0.34	0.33	0.35
B	0.26	0.35	0.07	0.33	0.35	0.34	0.33	0.31	0.37	0.32	0.35	0.33
C	0.32	0.28	0.18	0.14	0.33	0.26	0.32	0.29	0.34	0.28	0.32	0.28
D	0.01	0.29	0.31	0.36	0.27	0.33	0.35	0.34	0.32	0.34	0.35	0.03
E	0.34	0.31	0.31	0.14	0.33	0.34	0.08	0.30	0.31	0.32	0.35	0.31
F	0.30	0.31	0.38	0.34	0.32	0.36	0.35	0.32	0.33	0.34	0.33	0.32
G	0.35	0.31	0.33	0.33	0.32	0.33	0.33	0.31	0.34	0.30	0.27	0.33
H	0.33	0.31	0.35	0.22	0.32	0.05	0.35	0.30	0.36	0.04	0.07	0.06

R/S

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.81	0.82	0.82	0.95	0.93	1.02	1.03	0.88	0.91	1.00	0.98	0.94
B	0.88	0.86	-	0.96	0.95	0.93	0.95	0.88	0.86	1.03	0.88	0.96
C	1.03	1.09	-	-	0.92	1.06	0.87	0.85	0.87	0.97	0.90	1.16
D	-	0.98	0.99	0.86	0.61	0.94	0.83	0.84	1.01	0.98	0.79	-
E	0.99	0.88	0.92	-	0.89	0.97	-	0.44	0.97	0.71	0.93	1.12
F	0.98	0.68	0.87	0.89	1.00	0.88	0.87	0.82	0.6	0.81	1.03	1.12
G	0.88	0.98	0.97	0.97	0.61	0.91	1.00	1.02	0.93	1.05	0.29	1.05
H	1.01	0.91	0.94	1.28	1.05	-	0.78	1.14	0.93	-	-	-

Figure A.10: As Figure A.8 for library plates 5 (left) and 6 (right).

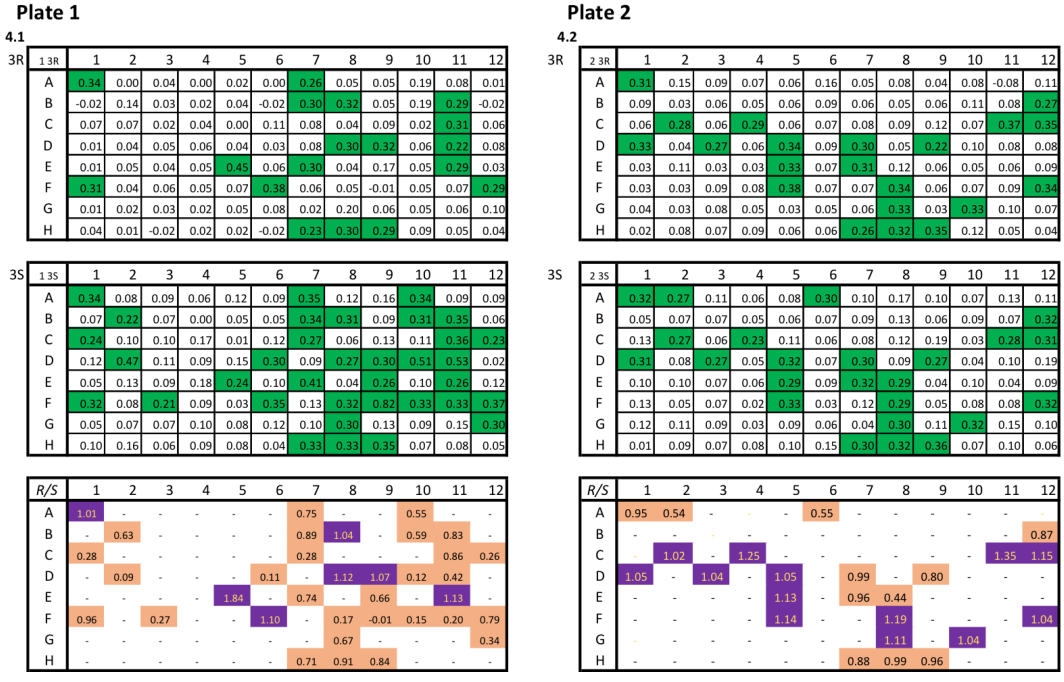


Plate 1

4.1 — 3R

1 3R	1	2	3	4	5	6	7	8	9	10	11	12
A	0.34	0.00	0.04	0.00	0.02	0.00	0.26	0.05	0.05	0.19	0.08	0.01
B	-0.02	0.14	0.03	0.02	0.04	-0.02	0.30	0.32	0.05	0.19	0.29	-0.02
C	0.07	0.07	0.02	0.04	0.00	0.11	0.08	0.04	0.09	0.02	0.31	0.06
D	0.01	0.04	0.05	0.06	0.04	0.03	0.08	0.30	0.32	0.06	0.22	0.08
E	0.01	0.05	0.04	0.05	0.45	0.06	0.30	0.04	0.17	0.05	0.29	0.03
F	0.31	0.04	0.06	0.05	0.07	0.38	0.06	0.05	-0.01	0.05	0.07	0.29
G	0.01	0.02	0.03	0.02	0.05	0.08	0.02	0.20	0.06	0.05	0.06	0.10
H	0.04	0.01	-0.02	0.02	0.02	-0.02	0.23	0.30	0.29	0.09	0.05	0.04

3S

1 3S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.34	0.08	0.09	0.06	0.12	0.09	0.35	0.12	0.16	0.34	0.09	0.09
B	0.07	0.22	0.07	0.00	0.05	0.05	0.34	0.31	0.09	0.31	0.35	0.06
C	0.24	0.10	0.10	0.17	0.01	0.12	0.27	0.06	0.13	0.11	0.36	0.23
D	0.12	0.47	0.11	0.09	0.15	0.30	0.09	0.27	0.30	0.51	0.53	0.02
E	0.05	0.13	0.09	0.18	0.24	0.10	0.41	0.04	0.26	0.10	0.26	0.12
F	0.32	0.08	0.21	0.09	0.03	0.35	0.13	0.32	0.82	0.33	0.33	0.37
G	0.05	0.07	0.07	0.10	0.08	0.12	0.10	0.30	0.13	0.09	0.15	0.30
H	0.10	0.16	0.06	0.09	0.08	0.04	0.33	0.33	0.35	0.07	0.08	0.05

R/S

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	1.01	-	-	-	-	-	0.75	-	-	0.55	-	-
B	-	0.63	-	-	-	-	0.89	1.04	-	0.59	0.83	-
C	0.28	-	-	-	-	-	0.28	-	-	-	0.86	0.26
D	-	0.09	-	-	-	0.11	-	1.12	1.07	0.12	0.42	-
E	-	-	-	-	1.84	-	0.74	-	0.66	-	1.13	-
F	0.96	-	0.27	-	-	1.10	-	0.17	-0.01	0.15	0.20	0.79
G	-	-	-	-	-	-	-	0.67	-	-	-	0.34
H	-	-	-	-	-	-	0.71	0.91	0.84	-	-	-

Plate 2

4.2 — 3R

2 3R	1	2	3	4	5	6	7	8	9	10	11	12
A	0.31	0.15	0.09	0.07	0.06	0.16	0.05	0.08	0.04	0.08	-0.08	0.11
B	0.09	0.03	0.06	0.05	0.06	0.09	0.06	0.05	0.06	0.11	0.08	0.27
C	0.06	0.28	0.06	0.29	0.06	0.07	0.08	0.09	0.12	0.07	0.37	0.35
D	0.33	0.04	0.27	0.06	0.34	0.09	0.30	0.05	0.22	0.10	0.08	0.08
E	0.03	0.11	0.03	0.03	0.33	0.07	0.31	0.12	0.06	0.05	0.06	0.09
F	0.03	0.03	0.09	0.08	0.38	0.07	0.07	0.34	0.06	0.07	0.09	0.34
G	0.04	0.03	0.08	0.05	0.03	0.05	0.06	0.33	0.03	0.33	0.10	0.07
H	0.02	0.08	0.07	0.09	0.06	0.06	0.26	0.32	0.35	0.12	0.05	0.04

3S

2 3S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.32	0.27	0.11	0.06	0.08	0.30	0.10	0.17	0.10	0.07	0.13	0.11
B	0.05	0.07	0.07	0.05	0.06	0.07	0.09	0.13	0.06	0.09	0.07	0.32
C	0.13	0.27	0.06	0.23	0.11	0.06	0.08	0.12	0.19	0.03	0.28	0.31
D	0.31	0.08	0.27	0.05	0.32	0.07	0.30	0.09	0.27	0.04	0.10	0.19
E	0.10	0.10	0.07	0.06	0.29	0.09	0.32	0.29	0.04	0.10	0.04	0.09
F	0.13	0.05	0.07	0.02	0.33	0.03	0.12	0.29	0.05	0.08	0.08	0.32
G	0.12	0.11	0.09	0.03	0.09	0.06	0.04	0.30	0.11	0.32	0.15	0.10
H	0.01	0.09	0.07	0.08	0.10	0.15	0.30	0.32	0.36	0.07	0.10	0.06

R/S

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.95	0.54	-	-	-	0.55	-	-	-	-	-	-
B	-	-	-	-	-	-	-	-	-	-	-	0.87
C	-	1.02	-	1.25	-	-	-	-	-	-	1.35	1.15
D	1.05	-	1.04	-	1.05	-	0.99	-	0.80	-	-	-
E	-	-	-	-	1.13	-	0.96	0.44	-	-	-	-
F	-	-	-	-	1.14	-	-	1.19	-	-	-	1.04
G	-	-	-	-	-	-	-	1.11	-	1.04	-	-
H	-	-	-	-	-	-	0.88	0.99	0.96	-	-	-

Figure A.11: Screening plates for CASTing group 4 ($Y30_{NDT}$ $R229_{NDT}$) library plates 1 (left) and 2 (right). For each library plate containing eTA variants, two screening plates are set up identically to measure retro-aldol activity of either 2*S*3*R* (top) or 2*S*3*S* (middle) PS. Positive controls are shown in wells H7-9 and negative controls in wells H10-12. The activity of each well is measured from the change in Abs_{260}. Those wells with an activity of 0.2 ΔAbs_{260} or greater are highlighted in green. Identifier analysis (bottom) shows the 2*S*3*R* / 2*S*3*S* activity ratio for wells highlighted in green above. Those in purple will have greater levels of activity for the 2*S*3*R* PS retro-aldol cleavage and those in brown for 2*S*3*S*.

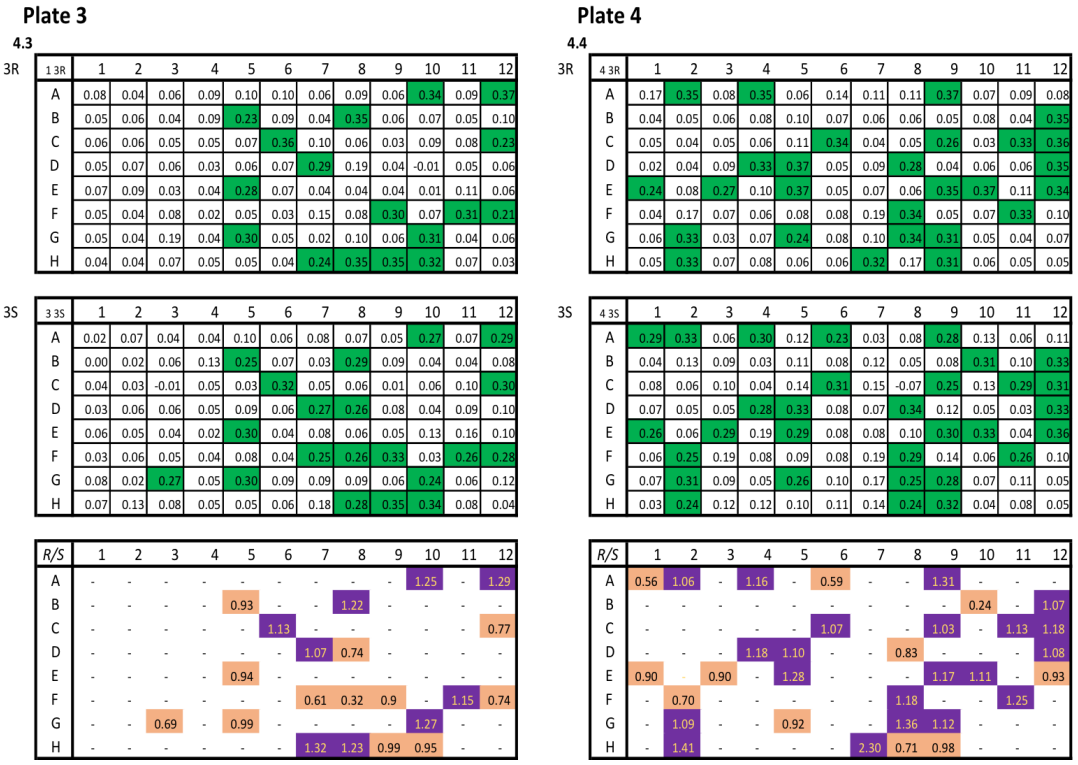


Figure A.12: As Figure A.11 for library plates 3 (left) and 4 (right).

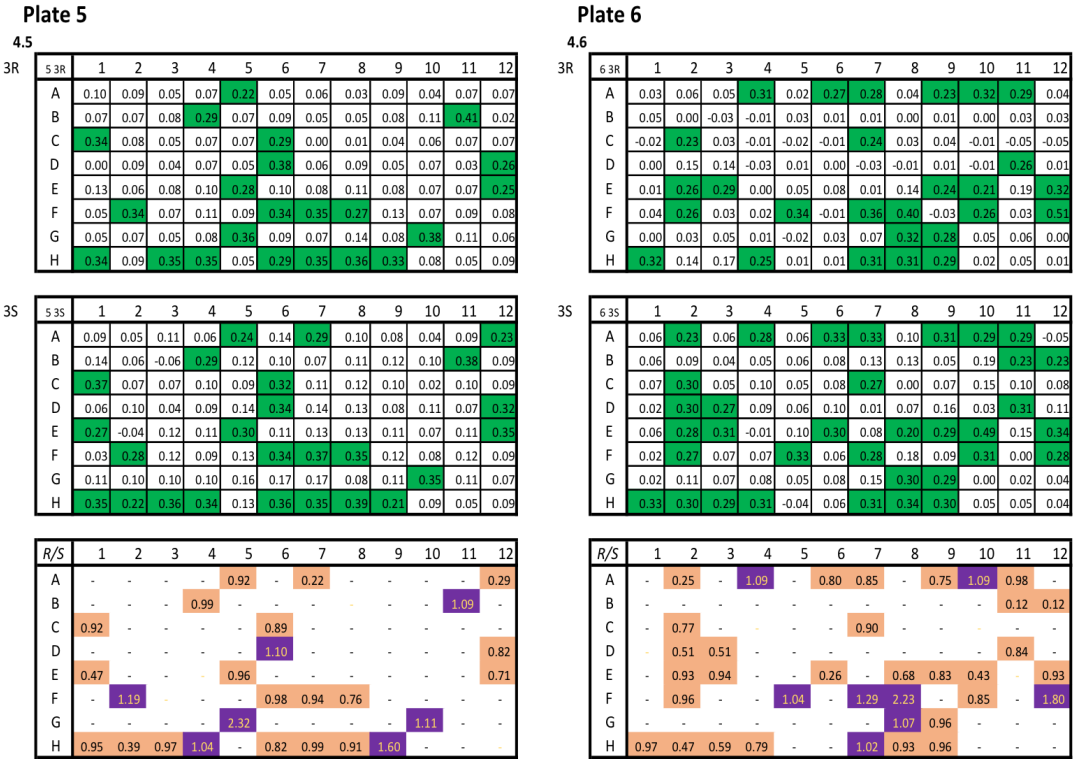


Figure A.13: As Figure A.11 for library plates 5 (left) and 6 (right).

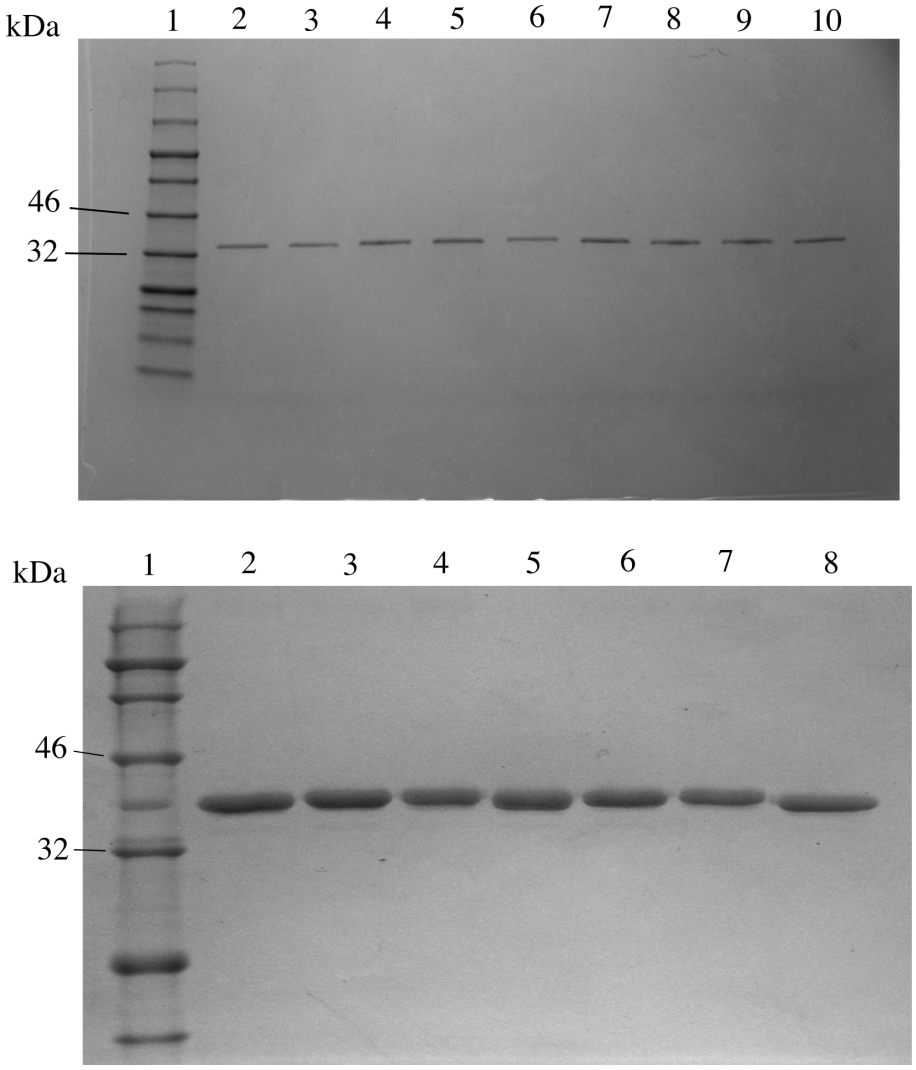


Figure A.14: SDS PAGE analysis of Group 4 (Y30F$_{NDT}$R229N$_{NDT}$) purified variants. Top: Lanes: 1 - protein ladder, 2 - eTA, 3 - 1 A10, 4 - 1 C1, 5 - 1 C12, 6 - 1 D11, 7 - 1 F11, 8 - 2 A2, 9 - 2 A6, 10 - 2 D12. Bottom: Lanes: 1 - protein ladder, 2 - eTA, 3 - 2 H6, 4 - 3 F8, 5 - 4 B10, 6 - 5 E1, 7 - 6 A2, 8 - 6 F8.

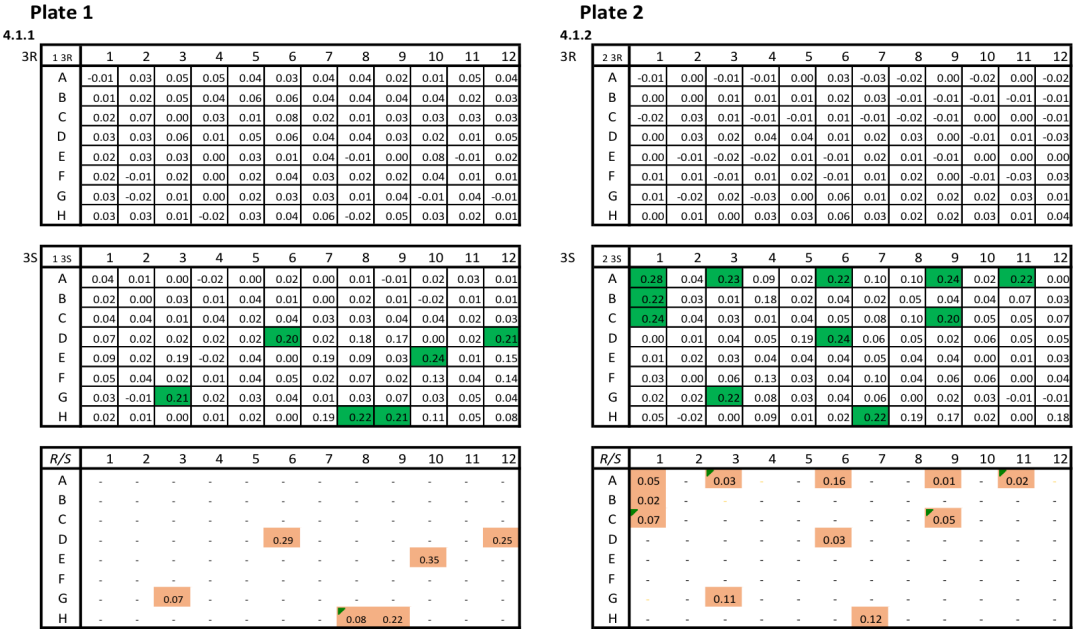


Plate 1

4.1.1

3R	1	2	3	4	5	6	7	8	9	10	11	12
A	-0.01	0.03	0.05	0.05	0.04	0.03	0.04	0.04	0.02	0.01	0.05	0.04
B	0.01	0.02	0.05	0.04	0.06	0.06	0.04	0.04	0.04	0.04	0.02	0.03
C	0.02	0.07	0.00	0.03	0.01	0.08	0.02	0.01	0.03	0.03	0.03	0.03
D	0.03	0.03	0.06	0.01	0.05	0.06	0.04	0.04	0.03	0.02	0.01	0.05
E	0.02	0.03	0.03	0.00	0.03	0.01	0.04	-0.01	0.00	0.08	-0.01	0.02
F	0.02	-0.01	0.02	0.00	0.02	0.04	0.03	0.02	0.02	0.04	0.01	0.01
G	0.03	-0.02	0.01	0.00	0.02	0.03	0.03	0.01	0.04	-0.01	0.04	-0.01
H	0.03	0.03	0.01	-0.02	0.03	0.04	0.06	-0.02	0.05	0.03	0.02	0.01

3S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.04	0.01	0.00	-0.02	0.00	0.02	0.00	0.01	-0.01	0.02	0.03	0.01
B	0.02	0.00	0.03	0.01	0.04	0.01	0.00	0.02	0.01	-0.02	0.01	0.01
C	0.04	0.04	0.01	0.04	0.02	0.04	0.03	0.03	0.04	0.04	0.02	0.03
D	0.07	0.02	0.02	0.02	0.02	0.20	0.02	0.18	0.17	0.00	0.02	0.21
E	0.09	0.02	0.19	-0.02	0.04	0.00	0.19	0.09	0.03	0.24	0.01	0.15
F	0.05	0.04	0.02	0.01	0.04	0.05	0.02	0.07	0.02	0.13	0.04	0.14
G	0.03	-0.01	0.21	0.02	0.03	0.04	0.01	0.03	0.07	0.03	0.05	0.04
H	0.02	0.01	0.00	0.01	0.02	0.00	0.19	0.22	0.21	0.11	0.05	0.08

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	-	-	-	-	-	-	-	-	-	-
B	-	-	-	-	-	-	-	-	-	-	-	-
C	-	-	-	-	-	-	-	-	-	-	-	-
D	-	-	-	-	-	0.29	-	-	-	-	-	0.25
E	-	-	-	-	-	-	-	-	-	0.35	-	-
F	-	-	-	-	-	-	-	-	-	-	-	-
G	-	-	0.07	-	-	-	-	-	-	-	-	-
H	-	-	-	-	-	-	-	0.08	0.22	-	-	-

Plate 2

4.1.2

3R	1	2	3	4	5	6	7	8	9	10	11	12
A	-0.01	0.00	-0.01	-0.01	0.00	0.03	-0.03	-0.02	0.00	-0.02	0.00	-0.02
B	0.00	0.00	0.01	0.01	0.01	0.02	0.03	-0.01	-0.01	-0.01	-0.01	-0.01
C	-0.02	0.03	0.01	-0.01	-0.01	0.01	-0.01	-0.02	-0.01	0.00	0.00	-0.01
D	0.00	0.03	0.02	0.04	0.04	0.01	0.02	0.03	0.00	-0.01	0.01	-0.03
E	0.00	-0.01	-0.02	-0.02	0.01	-0.01	0.02	0.01	-0.01	0.00	0.00	0.00
F	0.01	0.01	-0.01	0.01	0.02	-0.01	0.01	0.02	0.00	-0.01	-0.03	0.03
G	0.01	-0.02	0.02	-0.03	0.00	0.06	0.01	0.02	0.02	0.02	0.03	0.01
H	0.00	0.01	0.00	0.03	0.03	0.06	0.03	0.02	0.02	0.03	0.01	0.04

3S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.28	0.04	0.23	0.09	0.02	0.22	0.10	0.10	0.24	0.02	0.22	0.00
B	0.22	0.03	0.01	0.18	0.02	0.04	0.02	0.05	0.04	0.04	0.07	0.03
C	0.24	0.04	0.03	0.01	0.04	0.05	0.08	0.10	0.20	0.05	0.05	0.07
D	0.00	0.01	0.04	0.05	0.19	0.24	0.06	0.05	0.02	0.06	0.05	0.05
E	0.01	0.02	0.03	0.04	0.04	0.04	0.05	0.04	0.04	0.00	0.01	0.03
F	0.03	0.00	0.06	0.13	0.03	0.04	0.10	0.04	0.06	0.06	0.00	0.04
G	0.02	0.02	0.22	0.08	0.03	0.04	0.06	0.00	0.02	0.03	-0.01	-0.01
H	0.05	-0.02	0.00	0.09	0.01	0.02	0.22	0.19	0.17	0.02	0.00	0.18

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.05	-	0.03	-	-	0.16	-	-	0.01	-	0.02	-
B	0.02	-	-	-	-	-	-	-	-	-	-	-
C	0.07	-	-	-	-	-	-	-	0.05	-	-	-
D	-	-	-	-	-	0.03	-	-	-	-	-	-
E	-	-	-	-	-	-	-	-	-	-	-	-
F	-	-	-	-	-	-	-	-	-	-	-	-
G	-	-	0.11	-	-	-	-	-	-	-	-	-
H	-	-	-	-	-	-	0.12	-	-	-	-	-

Figure A.15: Screening plates for CASTing Round 2 group 1 (Y30F R229N + H83$_{NDT}$ F87$_{NDT}$) library plates 1 (left) and 2 (right). For each library plate containing eTA variants, two screening plates are set up identically to measure retro-aldol activity of either 2*S*3*R* (top) or 2*S*3*S* (middle) PS. Positive controls are shown in wells H7-9 and negative controls in wells H10-12. The activity of each well is measured from the change in Abs$_{260}$. Those wells with an activity of 0.2 ΔAbs$_{260}$ or greater are highlighted in green. Identifier analysis (bottom) shows the 2*S*3*R* / 2*S*3*S* activity ratio for wells highlighted in green above. Those in purple will have greater levels of activity for the 2*S*3*R* PS retro-aldol cleavage and those in brown for 2*S*3*S*.

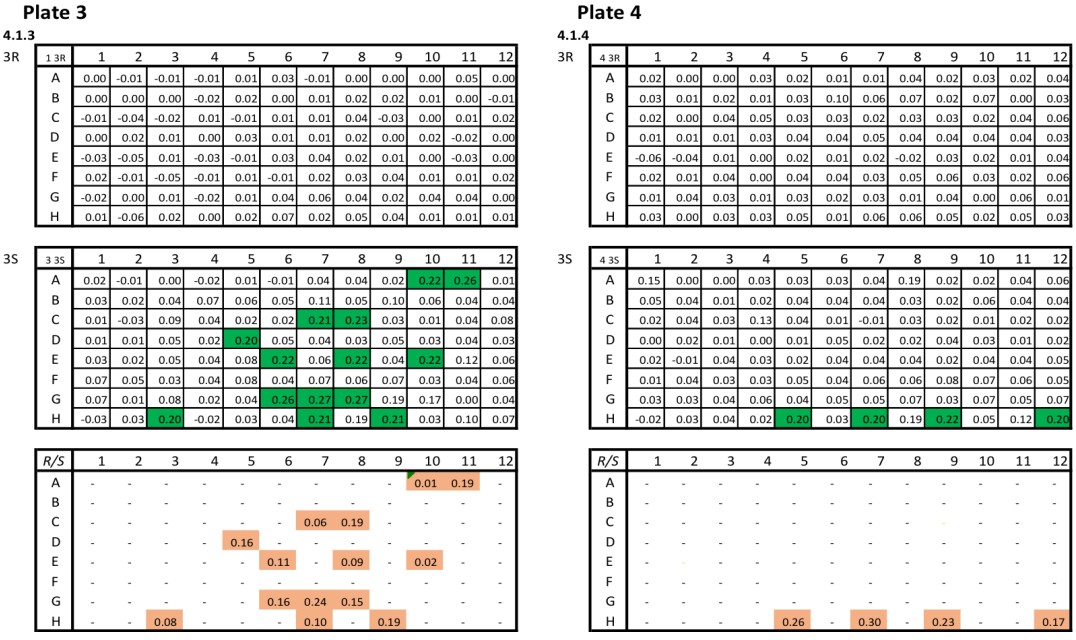


Plate 3

4.1.3

3R	1	2	3	4	5	6	7	8	9	10	11	12
A	0.00	-0.01	-0.01	-0.01	0.01	0.03	-0.01	0.00	0.00	0.00	0.05	0.00
B	0.00	0.00	0.00	-0.02	0.02	0.00	0.01	0.02	0.02	0.01	0.00	-0.01
C	-0.01	-0.04	-0.02	0.01	-0.01	0.01	0.01	0.04	-0.03	0.00	0.01	0.02
D	0.00	0.02	0.01	0.00	0.03	0.01	0.01	0.02	0.00	0.02	-0.02	0.00
E	-0.03	-0.05	0.01	-0.03	-0.01	0.03	0.04	0.02	0.01	0.00	-0.03	0.00
F	0.02	-0.01	-0.05	-0.01	0.01	-0.01	0.02	0.03	0.04	0.01	0.01	0.02
G	-0.02	0.00	0.01	-0.02	0.01	0.04	0.06	0.04	0.02	0.04	0.04	0.00
H	0.01	-0.06	0.02	0.00	0.02	0.07	0.02	0.05	0.04	0.01	0.01	0.01

3S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.02	-0.01	0.00	-0.02	0.01	-0.01	0.04	0.04	0.02	0.22	0.26	0.01
B	0.03	0.02	0.04	0.07	0.06	0.05	0.11	0.05	0.10	0.06	0.04	0.04
C	0.01	-0.03	0.09	0.04	0.02	0.02	0.21	0.23	0.03	0.01	0.04	0.08
D	0.01	0.01	0.05	0.02	0.20	0.05	0.04	0.03	0.05	0.03	0.04	0.03
E	0.03	0.02	0.05	0.04	0.08	0.22	0.06	0.22	0.04	0.22	0.12	0.06
F	0.07	0.05	0.03	0.04	0.08	0.04	0.07	0.06	0.07	0.03	0.04	0.06
G	0.07	0.01	0.08	0.02	0.04	0.26	0.27	0.27	0.19	0.17	0.00	0.04
H	-0.03	0.03	0.20	-0.02	0.03	0.04	0.21	0.19	0.21	0.03	0.10	0.07

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	-	-	-	-	-	-	-	0.01	0.19	-
B	-	-	-	-	-	-	-	-	-	-	-	-
C	-	-	-	-	-	0.06	0.19	-	-	-	-	-
D	-	-	-	-	0.16	-	-	-	-	-	-	-
E	-	-	-	-	-	0.11	-	0.09	-	0.02	-	-
F	-	-	-	-	-	-	-	-	-	-	-	-
G	-	-	-	-	-	0.16	0.24	0.15	-	-	-	-
H	-	-	0.08	-	-	-	0.10	-	0.19	-	-	-

Plate 4

4.1.4

3R	1	2	3	4	5	6	7	8	9	10	11	12
A	0.02	0.00	0.00	0.03	0.02	0.01	0.01	0.04	0.02	0.03	0.02	0.04
B	0.03	0.01	0.02	0.01	0.03	0.10	0.06	0.07	0.02	0.07	0.00	0.03
C	0.02	0.00	0.04	0.05	0.03	0.03	0.02	0.03	0.03	0.02	0.04	0.06
D	0.01	0.01	0.01	0.03	0.04	0.04	0.05	0.04	0.04	0.04	0.04	0.03
E	-0.06	-0.04	0.01	0.00	0.02	0.01	0.02	-0.02	0.03	0.02	0.01	0.04
F	0.02	0.01	0.04	0.00	0.04	0.04	0.03	0.05	0.06	0.03	0.02	0.06
G	0.01	0.04	0.03	0.01	0.03	0.02	0.03	0.01	0.04	0.00	0.06	0.01
H	0.03	0.00	0.03	0.03	0.05	0.01	0.06	0.06	0.05	0.02	0.05	0.03

3S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.15	0.00	0.00	0.03	0.03	0.03	0.04	0.19	0.02	0.02	0.04	0.06
B	0.05	0.04	0.01	0.02	0.04	0.04	0.04	0.03	0.02	0.06	0.04	0.04
C	0.02	0.04	0.03	0.13	0.04	0.01	-0.01	0.03	0.02	0.01	0.02	0.02
D	0.00	0.02	0.01	0.00	0.01	0.05	0.02	0.02	0.04	0.03	0.01	0.02
E	0.02	-0.01	0.04	0.03	0.02	0.04	0.04	0.04	0.02	0.04	0.04	0.05
F	0.01	0.04	0.03	0.03	0.05	0.04	0.06	0.06	0.08	0.07	0.06	0.05
G	0.03	0.03	0.04	0.06	0.04	0.05	0.05	0.07	0.03	0.07	0.05	0.07
H	-0.02	0.03	0.04	0.02	0.20	0.03	0.20	0.19	0.22	0.05	0.12	0.20

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	-	-	-	-	-	-	-	-	-	-
B	-	-	-	-	-	-	-	-	-	-	-	-
C	-	-	-	-	-	-	-	-	-	-	-	-
D	-	-	-	-	-	-	-	-	-	-	-	-
E	-	-	-	-	-	-	-	-	-	-	-	-
F	-	-	-	-	-	-	-	-	-	-	-	-
G	-	-	-	-	-	-	-	-	-	-	-	-
H	-	-	-	-	0.26	-	0.30	-	0.23	-	-	0.17

Figure A.16: As Figure A.15 for library plates 3 (left) and 4 (right).

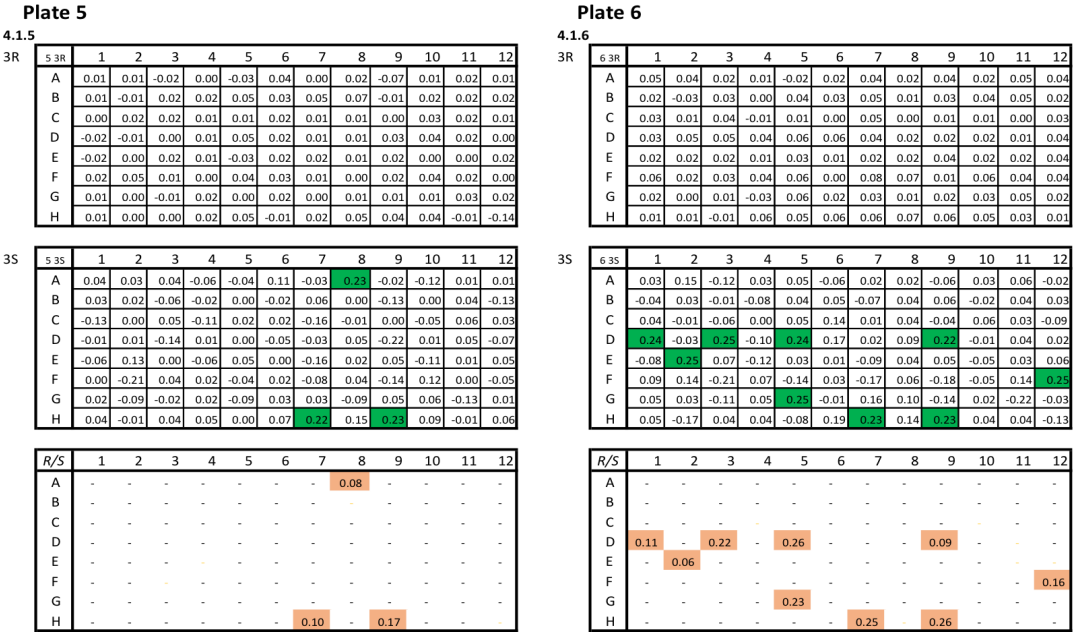


Plate 5

4.1.5

3R

5 3R	1	2	3	4	5	6	7	8	9	10	11	12
A	0.01	0.01	-0.02	0.00	-0.03	0.04	0.00	0.02	-0.07	0.01	0.02	0.01
B	0.01	-0.01	0.02	0.02	0.05	0.03	0.05	0.07	-0.01	0.02	0.02	0.02
C	0.00	0.02	0.02	0.01	0.01	0.02	0.01	0.01	0.00	0.03	0.02	0.01
D	-0.02	-0.01	0.00	0.01	0.05	0.02	0.01	0.01	0.03	0.04	0.02	0.00
E	-0.02	0.00	0.02	0.01	-0.03	0.02	0.02	0.01	0.02	0.00	0.00	0.02
F	0.02	0.05	0.01	0.00	0.04	0.03	0.01	0.00	0.02	0.04	0.02	0.00
G	0.01	0.00	-0.01	0.02	0.00	0.02	0.00	0.01	0.01	0.01	0.03	0.02
H	0.01	0.00	0.00	0.02	0.05	-0.01	0.02	0.05	0.04	0.04	-0.01	-0.14

3S

5 3S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.04	0.03	0.04	-0.06	-0.04	0.11	-0.03	0.23	-0.02	-0.12	0.01	0.01
B	0.03	0.02	-0.06	-0.02	0.00	-0.02	0.06	0.00	-0.13	0.00	0.04	-0.13
C	-0.13	0.00	0.05	-0.11	0.02	0.02	-0.16	-0.01	0.00	-0.05	0.06	0.03
D	-0.01	0.01	-0.14	0.01	0.00	-0.05	-0.03	0.05	-0.22	0.01	0.05	-0.07
E	-0.06	0.13	0.00	-0.06	0.05	0.00	-0.16	0.02	0.05	-0.11	0.01	0.05
F	0.00	-0.21	0.04	0.02	-0.04	0.02	-0.08	0.04	-0.14	0.12	0.00	-0.05
G	0.02	-0.09	-0.02	0.02	-0.09	0.03	0.03	-0.09	0.05	0.06	-0.13	0.01
H	0.04	-0.01	0.04	0.05	0.00	0.07	0.22	0.15	0.23	0.09	-0.01	0.06

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	-	-	-	-	-	0.08	-	-	-	-
B	-	-	-	-	-	-	-	-	-	-	-	-
C	-	-	-	-	-	-	-	-	-	-	-	-
D	-	-	-	-	-	-	-	-	-	-	-	-
E	-	-	-	-	-	-	-	-	-	-	-	-
F	-	-	-	-	-	-	-	-	-	-	-	-
G	-	-	-	-	-	-	-	-	-	-	-	-
H	-	-	-	-	-	-	0.10	-	0.17	-	-	-

Plate 6

4.1.6

3R

6 3R	1	2	3	4	5	6	7	8	9	10	11	12
A	0.05	0.04	0.02	0.01	-0.02	0.02	0.04	0.02	0.04	0.02	0.05	0.04
B	0.02	-0.03	0.03	0.00	0.04	0.03	0.05	0.01	0.03	0.04	0.05	0.02
C	0.03	0.01	0.04	-0.01	0.01	0.00	0.05	0.00	0.01	0.01	0.00	0.03
D	0.03	0.05	0.05	0.04	0.06	0.06	0.04	0.02	0.02	0.02	0.01	0.04
E	0.02	0.02	0.02	0.01	0.03	0.01	0.02	0.02	0.04	0.02	0.02	0.04
F	0.06	0.02	0.03	0.04	0.06	0.00	0.08	0.07	0.01	0.06	0.04	0.04
G	0.02	0.00	0.01	-0.03	0.06	0.02	0.03	0.01	0.02	0.03	0.05	0.02
H	0.01	0.01	-0.01	0.06	0.05	0.06	0.06	0.07	0.06	0.05	0.03	0.01

3S

6 3S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.03	0.15	-0.12	0.03	0.05	-0.06	0.02	0.02	-0.06	0.03	0.06	-0.02
B	-0.04	0.03	-0.01	-0.08	0.04	0.05	-0.07	0.04	0.06	-0.02	0.04	0.03
C	0.04	-0.01	-0.06	0.00	0.05	0.14	0.01	0.04	-0.04	0.06	0.03	-0.09
D	0.24	-0.03	0.25	-0.10	0.24	0.17	0.02	0.09	0.22	-0.01	0.04	0.02
E	-0.08	0.25	0.07	-0.12	0.03	0.01	-0.09	0.04	0.05	-0.05	0.03	0.06
F	0.09	0.14	-0.21	0.07	-0.14	0.03	-0.17	0.06	-0.18	-0.05	0.14	0.25
G	0.05	0.03	-0.11	0.05	0.25	-0.01	0.16	0.10	-0.14	0.02	-0.22	-0.03
H	0.05	-0.17	0.04	0.04	-0.08	0.19	0.23	0.14	0.23	0.04	0.04	-0.13

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	-	-	-	-	-	-	-	-	-	-
B	-	-	-	-	-	-	-	-	-	-	-	-
C	-	-	-	-	-	-	-	-	-	-	-	-
D	0.11	-	0.22	-	0.26	-	-	-	0.09	-	-	-
E	-	0.06	-	-	-	-	-	-	-	-	-	-
F	-	-	-	-	-	-	-	-	-	-	-	0.16
G	-	-	-	-	0.23	-	-	-	-	-	-	-
H	-	-	-	-	-	-	0.25	-	0.26	-	-	-

Figure A.17: As Figure A.15 for library plates 5 (left) and 6 (right).

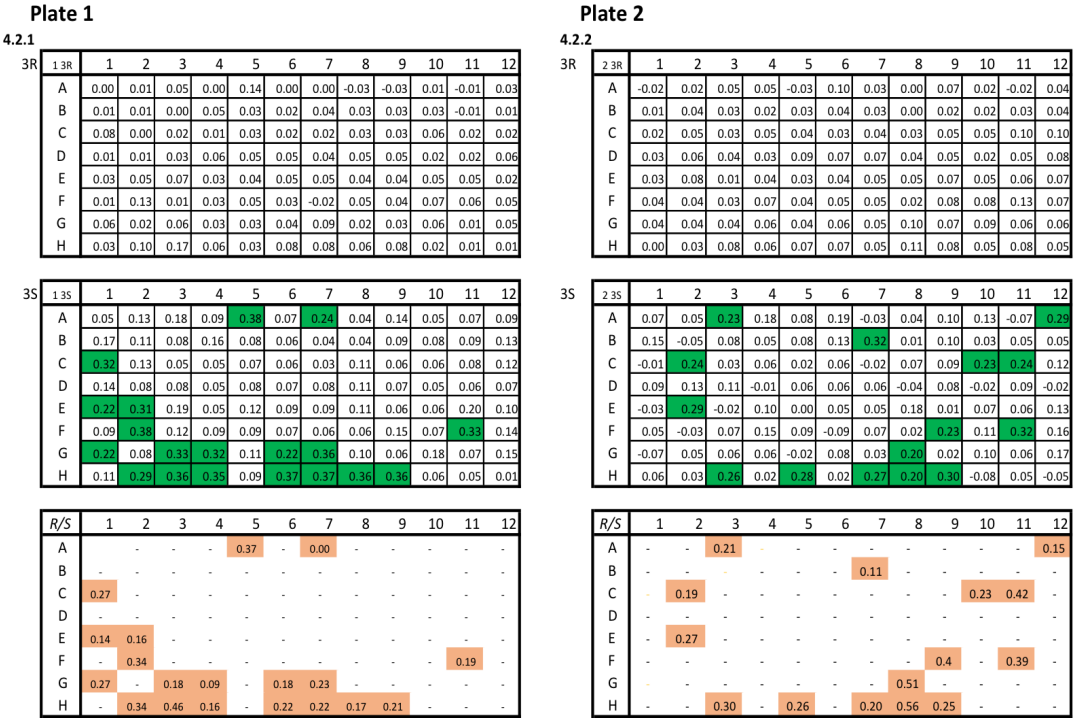


Plate 1

4.2.1

3R

1 3R	1	2	3	4	5	6	7	8	9	10	11	12
A	0.00	0.01	0.05	0.00	0.14	0.00	0.00	-0.03	-0.03	0.01	-0.01	0.03
B	0.01	0.01	0.00	0.05	0.03	0.02	0.04	0.03	0.03	0.03	-0.01	0.01
C	0.08	0.00	0.02	0.01	0.03	0.02	0.02	0.03	0.03	0.06	0.02	0.02
D	0.01	0.01	0.03	0.06	0.05	0.05	0.04	0.05	0.05	0.02	0.02	0.06
E	0.03	0.05	0.07	0.03	0.04	0.05	0.05	0.04	0.04	0.05	0.05	0.02
F	0.01	0.13	0.01	0.03	0.05	0.03	-0.02	0.05	0.04	0.07	0.06	0.05
G	0.06	0.02	0.06	0.03	0.03	0.04	0.09	0.02	0.03	0.06	0.01	0.05
H	0.03	0.10	0.17	0.06	0.03	0.08	0.08	0.06	0.08	0.02	0.01	0.01

3S

1 3S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.05	0.13	0.18	0.09	0.38	0.07	0.24	0.04	0.14	0.05	0.07	0.09
B	0.17	0.11	0.08	0.16	0.08	0.06	0.04	0.04	0.09	0.08	0.09	0.13
C	0.32	0.13	0.05	0.05	0.07	0.06	0.03	0.11	0.06	0.06	0.08	0.12
D	0.14	0.08	0.08	0.05	0.08	0.07	0.08	0.11	0.07	0.05	0.06	0.07
E	0.22	0.31	0.19	0.05	0.12	0.09	0.09	0.11	0.06	0.06	0.20	0.10
F	0.09	0.38	0.12	0.09	0.09	0.07	0.06	0.06	0.15	0.07	0.33	0.14
G	0.22	0.08	0.33	0.32	0.11	0.22	0.36	0.10	0.06	0.18	0.07	0.15
H	0.11	0.29	0.36	0.35	0.09	0.37	0.37	0.36	0.36	0.06	0.05	0.01

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	-	-	0.37	-	0.00	-	-	-	-	-
B	-	-	-	-	-	-	-	-	-	-	-	-
C	0.27	-	-	-	-	-	-	-	-	-	-	-
D	-	-	-	-	-	-	-	-	-	-	-	-
E	0.14	0.16	-	-	-	-	-	-	-	-	-	-
F	-	0.34	-	-	-	-	-	-	-	-	0.19	-
G	0.27	-	0.18	0.09	-	0.18	0.23	-	-	-	-	-
H	-	0.34	0.46	0.16	-	0.22	0.22	0.17	0.21	-	-	-

Plate 2

4.2.2

3R

2 3R	1	2	3	4	5	6	7	8	9	10	11	12
A	-0.02	0.02	0.05	0.05	-0.03	0.10	0.03	0.00	0.07	0.02	-0.02	0.04
B	0.01	0.04	0.03	0.02	0.03	0.04	0.03	0.00	0.02	0.02	0.03	0.04
C	0.02	0.05	0.03	0.05	0.04	0.03	0.04	0.03	0.05	0.05	0.10	0.10
D	0.03	0.06	0.04	0.03	0.09	0.07	0.07	0.04	0.05	0.02	0.05	0.08
E	0.03	0.08	0.01	0.04	0.03	0.04	0.05	0.05	0.07	0.05	0.06	0.07
F	0.04	0.04	0.03	0.07	0.04	0.05	0.05	0.02	0.08	0.08	0.13	0.07
G	0.04	0.04	0.04	0.06	0.04	0.06	0.05	0.10	0.07	0.09	0.06	0.06
H	0.00	0.03	0.08	0.06	0.07	0.07	0.05	0.11	0.08	0.05	0.08	0.05

3S

2 3S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.07	0.05	0.23	0.18	0.08	0.19	-0.03	0.04	0.10	0.13	-0.07	0.29
B	0.15	-0.05	0.08	0.05	0.08	0.13	0.32	0.01	0.10	0.03	0.05	0.05
C	-0.01	0.24	0.03	0.06	0.02	0.06	-0.02	0.07	0.09	0.23	0.24	0.12
D	0.09	0.13	0.11	-0.01	0.06	0.06	0.06	-0.04	0.08	-0.02	0.09	-0.02
E	-0.03	0.29	-0.02	0.10	0.00	0.05	0.05	0.18	0.01	0.07	0.06	0.13
F	0.05	-0.03	0.07	0.15	0.09	-0.09	0.07	0.02	0.23	0.11	0.32	0.16
G	-0.07	0.05	0.06	0.06	-0.02	0.08	0.03	0.20	0.02	0.10	0.06	0.17
H	0.06	0.03	0.26	0.02	0.28	0.02	0.27	0.20	0.30	-0.08	0.05	-0.05

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	0.21	-	-	-	-	-	-	-	-	0.15
B	-	-	-	-	-	-	0.11	-	-	-	-	-
C	-	0.19	-	-	-	-	-	-	-	0.23	0.42	-
D	-	-	-	-	-	-	-	-	-	-	-	-
E	-	0.27	-	-	-	-	-	-	-	-	-	-
F	-	-	-	-	-	-	-	-	0.4	-	0.39	-
G	-	-	-	-	-	-	-	0.51	-	-	-	-
H	-	-	0.30	-	0.26	-	0.20	0.56	0.25	-	-	-

Figure A.18: Screening plates for CASTing Round 2 group 2 (Y30F R229N + H126$_{NDT}$ F127$_{NDT}$) library plates 1 (left) and 2 (right). For each library plate containing eTA variants, two screening plates are set up identically to measure retro-aldol activity of either 2*S*3*R* (top) or 2*S*3*S* (middle) PS. Positive controls are shown in wells H7-9 and negative controls in wells H10-12. The activity of each well is measured from the change in Abs$_{260}$. Those wells with an activity of 0.2 ΔAbs$_{260}$ or greater are highlighted in green. Identifier analysis (bottom) shows the 2*S*3*R* / 2*S*3*S* activity ratio for wells highlighted in green above. Those in purple will have greater levels of activity for the 2*S*3*R* PS retro-aldol cleavage and those in brown for 2*S*3*S*.

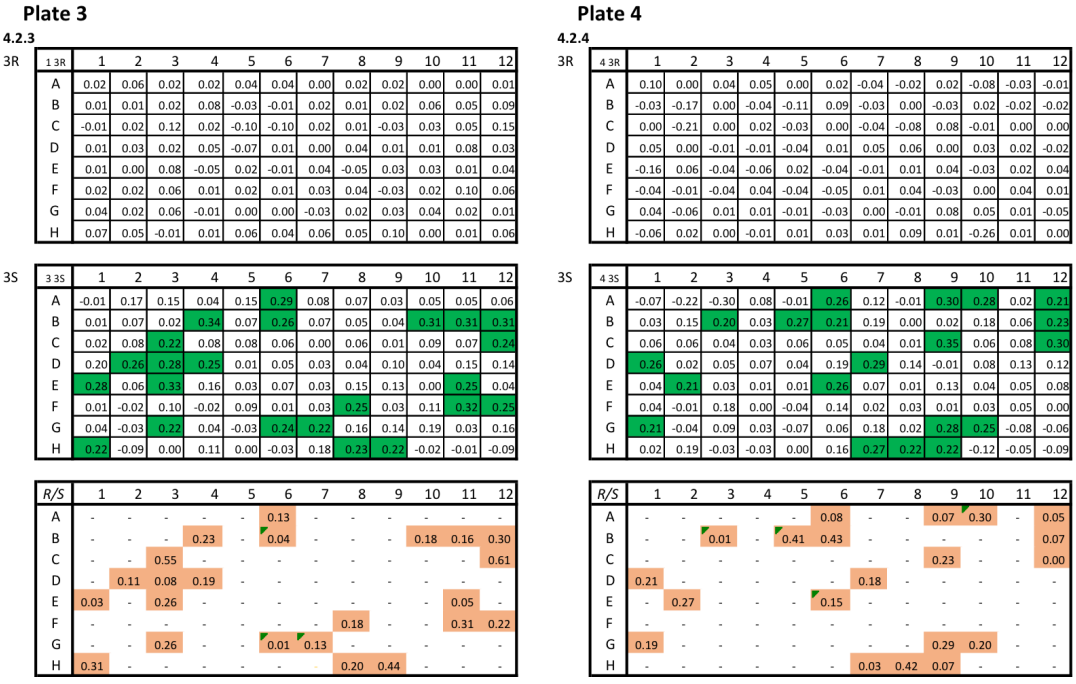


Plate 3

4.2.3

3R

1 3R	1	2	3	4	5	6	7	8	9	10	11	12
A	0.02	0.06	0.02	0.02	0.04	0.04	0.00	0.02	0.02	0.00	0.00	0.01
B	0.01	0.01	0.02	0.08	-0.03	-0.01	0.02	0.01	0.02	0.06	0.05	0.09
C	-0.01	0.02	0.12	0.02	-0.10	-0.10	0.02	0.01	-0.03	0.03	0.05	0.15
D	0.01	0.03	0.02	0.05	-0.07	0.01	0.00	0.04	0.01	0.01	0.08	0.03
E	0.01	0.00	0.08	-0.05	0.02	-0.01	0.04	-0.05	0.03	0.03	0.01	0.04
F	0.02	0.02	0.06	0.01	0.02	0.01	0.03	0.04	-0.03	0.02	0.10	0.06
G	0.04	0.02	0.06	-0.01	0.00	0.00	-0.03	0.02	0.03	0.04	0.02	0.01
H	0.07	0.05	-0.01	0.01	0.06	0.04	0.06	0.05	0.10	0.00	0.01	0.06

3S

3 3S	1	2	3	4	5	6	7	8	9	10	11	12
A	-0.01	0.17	0.15	0.04	0.15	0.29	0.08	0.07	0.03	0.05	0.05	0.06
B	0.01	0.07	0.02	0.34	0.07	0.26	0.07	0.05	0.04	0.31	0.31	0.31
C	0.02	0.08	0.22	0.08	0.08	0.06	0.00	0.06	0.01	0.09	0.07	0.24
D	0.20	0.26	0.28	0.25	0.01	0.05	0.03	0.04	0.10	0.04	0.15	0.14
E	0.28	0.06	0.33	0.16	0.03	0.07	0.03	0.15	0.13	0.00	0.25	0.04
F	0.01	-0.02	0.10	-0.02	0.09	0.01	0.03	0.25	0.03	0.11	0.32	0.25
G	0.04	-0.03	0.22	0.04	-0.03	0.24	0.22	0.16	0.14	0.19	0.03	0.16
H	0.22	-0.09	0.00	0.11	0.00	-0.03	0.18	0.23	0.22	-0.02	-0.01	-0.09

R/S

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	-	-	-	0.13	-	-	-	-	-	-
B	-	-	-	0.23	-	0.04	-	-	-	0.18	0.16	0.30
C	-	-	0.55	-	-	-	-	-	-	-	-	0.61
D	-	0.11	0.08	0.19	-	-	-	-	-	-	-	-
E	0.03	-	0.26	-	-	-	-	-	-	-	0.05	-
F	-	-	-	-	-	-	-	0.18	-	-	0.31	0.22
G	-	-	0.26	-	-	0.01	0.13	-	-	-	-	-
H	0.31	-	-	-	-	-	-	0.20	0.44	-	-	-

Plate 4

4.2.4

3R

4 3R	1	2	3	4	5	6	7	8	9	10	11	12
A	0.10	0.00	0.04	0.05	0.00	0.02	-0.04	-0.02	0.02	-0.08	-0.03	-0.01
B	-0.03	-0.17	0.00	-0.04	-0.11	0.09	-0.03	0.00	-0.03	0.02	-0.02	-0.02
C	0.00	-0.21	0.00	0.02	-0.03	0.00	-0.04	-0.08	0.08	-0.01	0.00	0.00
D	0.05	0.00	-0.01	-0.01	-0.04	0.01	0.05	0.06	0.00	0.03	0.02	-0.02
E	-0.16	0.06	-0.04	-0.06	0.02	-0.04	-0.01	0.01	0.04	-0.03	0.02	0.04
F	-0.04	-0.01	-0.04	0.04	-0.04	-0.05	0.01	0.04	-0.03	0.00	0.04	0.01
G	0.04	-0.06	0.01	0.01	-0.01	-0.03	0.00	-0.01	0.08	0.05	0.01	-0.05
H	-0.06	0.02	0.00	-0.01	0.01	0.03	0.01	0.09	0.01	-0.26	0.01	0.00

3S

4 3S	1	2	3	4	5	6	7	8	9	10	11	12
A	-0.07	-0.22	-0.30	0.08	-0.01	0.26	0.12	-0.01	0.30	0.28	0.02	0.21
B	0.03	0.15	0.20	0.03	0.27	0.21	0.19	0.00	0.02	0.18	0.06	0.23
C	0.06	0.06	0.04	0.03	0.06	0.05	0.04	0.01	0.35	0.06	0.08	0.30
D	0.26	0.02	0.05	0.07	0.04	0.19	0.29	0.14	-0.01	0.08	0.13	0.12
E	0.04	0.21	0.03	0.01	0.01	0.26	0.07	0.01	0.13	0.04	0.05	0.08
F	0.04	-0.01	0.18	0.00	-0.04	0.14	0.02	0.03	0.01	0.03	0.05	0.00
G	0.21	-0.04	0.09	0.03	-0.07	0.06	0.18	0.02	0.28	0.25	-0.08	-0.06
H	0.02	0.19	-0.03	-0.03	0.00	0.16	0.27	0.22	0.22	-0.12	-0.05	-0.09

R/S

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	-	-	-	0.08	-	-	0.07	0.30	-	0.05
B	-	-	0.01	-	0.41	0.43	-	-	-	-	-	0.07	
C	-	-	-	-	-	-	-	-	0.23	-	-	0.00
D	0.21	-	-	-	-	-	0.18	-	-	-	-	-
E	-	0.27	-	-	-	0.15	-	-	-	-	-	-
F	-	-	-	-	-	-	-	-	-	-	-	-
G	0.19	-	-	-	-	-	-	-	0.29	0.20	-	-
H	-	-	-	-	-	-	0.03	0.42	0.07	-	-	-

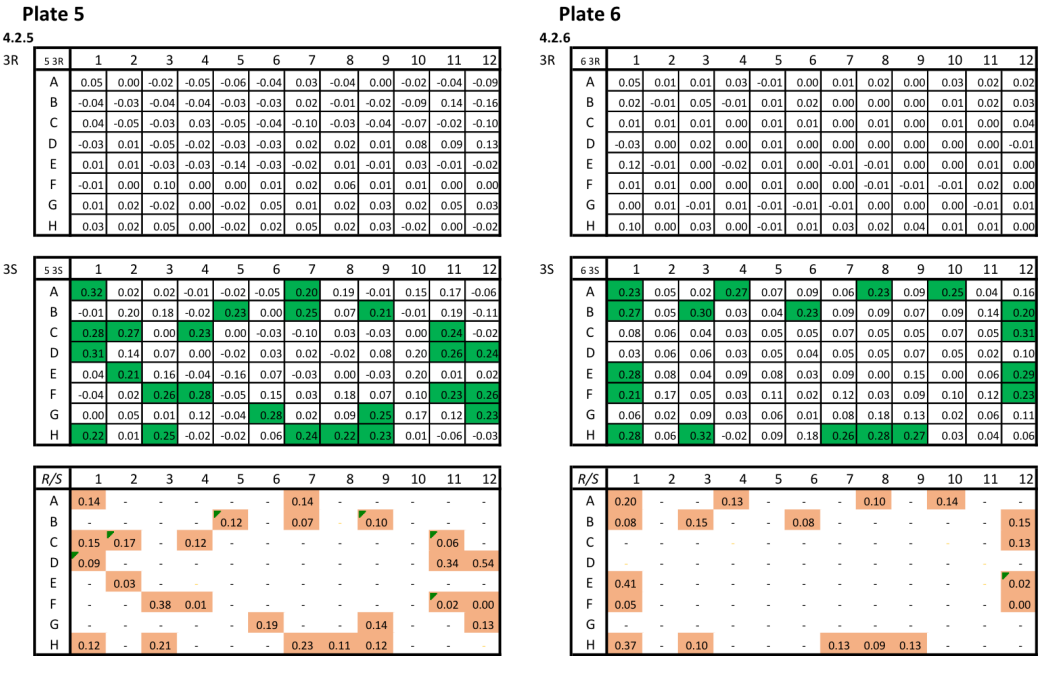


Plate 5

4.2.5

3R

5 3R	1	2	3	4	5	6	7	8	9	10	11	12
A	0.05	0.00	-0.02	-0.05	-0.06	-0.04	0.03	-0.04	0.00	-0.02	-0.04	-0.09
B	-0.04	-0.03	-0.04	-0.04	-0.03	-0.03	0.02	-0.01	-0.02	-0.09	0.14	-0.16
C	0.04	-0.05	-0.03	0.03	-0.05	-0.04	-0.10	-0.03	-0.04	-0.07	-0.02	-0.10
D	-0.03	0.01	-0.05	-0.02	-0.03	-0.03	0.02	0.02	0.01	0.08	0.09	0.13
E	0.01	0.01	-0.03	-0.03	-0.14	-0.03	-0.02	0.01	-0.01	0.03	-0.01	-0.02
F	-0.01	0.00	0.10	0.00	0.00	0.01	0.02	0.06	0.01	0.01	0.00	0.00
G	0.01	0.02	-0.02	0.00	-0.02	0.05	0.01	0.02	0.03	0.02	0.05	0.03
H	0.03	0.02	0.05	0.00	-0.02	0.02	0.05	0.02	0.03	-0.02	0.00	-0.02

3S

5 3S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.32	0.02	0.02	-0.01	-0.02	-0.05	0.20	0.19	-0.01	0.15	0.17	-0.06
B	-0.01	0.20	0.18	-0.02	0.23	0.00	0.25	0.07	0.21	-0.01	0.19	-0.11
C	0.28	0.27	0.00	0.23	0.00	-0.03	-0.10	0.03	-0.03	0.00	0.24	-0.02
D	0.31	0.14	0.07	0.00	-0.02	0.03	0.02	-0.02	0.08	0.20	0.26	0.24
E	0.04	0.21	0.16	-0.04	-0.16	0.07	-0.03	0.00	-0.03	0.20	0.01	0.02
F	-0.04	0.02	0.26	0.28	-0.05	0.15	0.03	0.18	0.07	0.10	0.23	0.26
G	0.00	0.05	0.01	0.12	-0.04	0.28	0.02	0.09	0.25	0.17	0.12	0.23
H	0.22	0.01	0.25	-0.02	-0.02	0.06	0.24	0.22	0.23	0.01	-0.06	-0.03

R/S

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.14	-	-	-	-	-	0.14	-	-	-	-	-
B	-	-	-	-	0.12	-	0.07	-	0.10	-	-	-
C	0.15	0.17	-	0.12	-	-	-	-	-	-	0.06	-
D	0.09	-	-	-	-	-	-	-	-	-	0.34	0.54
E	-	0.03	-	-	-	-	-	-	-	-	-	-
F	-	-	0.38	0.01	-	-	-	-	-	-	0.02	0.00
G	-	-	-	-	-	0.19	-	-	0.14	-	-	0.13
H	0.12	-	0.21	-	-	-	0.23	0.11	0.12	-	-	-

Plate 6

4.2.6

3R

6 3R	1	2	3	4	5	6	7	8	9	10	11	12
A	0.05	0.01	0.01	0.03	-0.01	0.00	0.01	0.02	0.00	0.03	0.02	0.02
B	0.02	-0.01	0.05	-0.01	0.01	0.02	0.00	0.00	0.00	0.01	0.02	0.03
C	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.00	0.04
D	-0.03	0.00	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	-0.01
E	0.12	-0.01	0.00	-0.02	0.01	0.00	-0.01	-0.01	0.00	0.00	0.01	0.00
F	0.01	0.01	0.00	0.00	0.01	0.00	0.00	-0.01	-0.01	-0.01	0.02	0.00
G	0.00	0.01	-0.01	0.01	-0.01	-0.01	-0.01	0.00	0.00	0.00	-0.01	0.01
H	0.10	0.00	0.03	0.00	-0.01	0.01	0.03	0.02	0.04	0.01	0.01	0.00

3S

6 3S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.23	0.05	0.02	0.27	0.07	0.09	0.06	0.23	0.09	0.25	0.04	0.16
B	0.27	0.05	0.30	0.03	0.04	0.23	0.09	0.09	0.07	0.09	0.14	0.20
C	0.08	0.06	0.04	0.03	0.05	0.05	0.07	0.05	0.05	0.07	0.05	0.31
D	0.03	0.06	0.06	0.03	0.05	0.04	0.05	0.05	0.07	0.05	0.02	0.10
E	0.28	0.08	0.04	0.09	0.08	0.03	0.09	0.00	0.15	0.00	0.06	0.29
F	0.21	0.17	0.05	0.03	0.11	0.02	0.12	0.03	0.09	0.10	0.12	0.23
G	0.06	0.02	0.09	0.03	0.06	0.01	0.08	0.18	0.13	0.02	0.06	0.11
H	0.28	0.06	0.32	-0.02	0.09	0.18	0.26	0.28	0.27	0.03	0.04	0.06

R/S

R/S	1	2	3	4	5	6	7	8	9	10	11	12
A	0.20	-	-	0.13	-	-	-	0.10	-	0.14	-	-
B	0.08	-	0.15	-	-	0.08	-	-	-	-	-	0.15
C	-	-	-	-	-	-	-	-	-	-	-	0.13
D	-	-	-	-	-	-	-	-	-	-	-	-
E	0.41	-	-	-	-	-	-	-	-	-	-	0.02
F	0.05	-	-	-	-	-	-	-	-	-	-	0.00
G	-	-	-	-	-	-	-	-	-	-	-	-
H	0.37	-	0.10	-	-	-	0.13	0.09	0.13	-	-	-

Figure A.19: As Figure A.18 for library plates 3 (left) and 4 (right).

Figure A.20: As Figure A.18 for library plates 5 (left) and 6 (right).

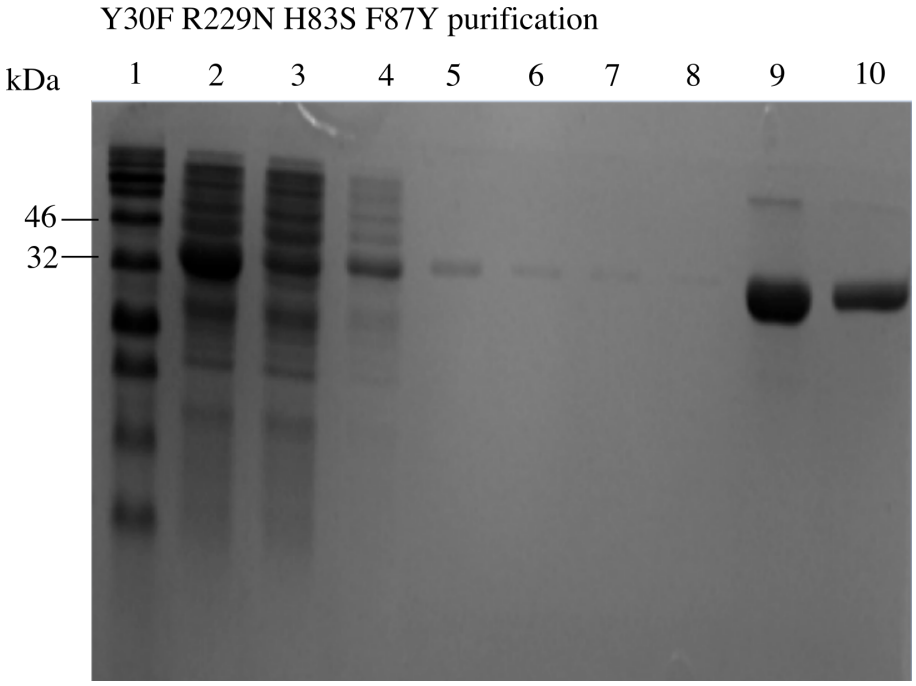


Figure A.21: SDS PAGE analysis of fractions from the purification for Y30F R229N H83S F87Y eTA. Lanes: 1 - protein ladder, 2 - soluble, 3 - insoluble, 4 - unbound, 5 - wash 1, 6 - wash 2, 7 - wash 3, 8 - wash 4, 9 - elution at 20 times dilute, 10 - elution at 40 times dilute.

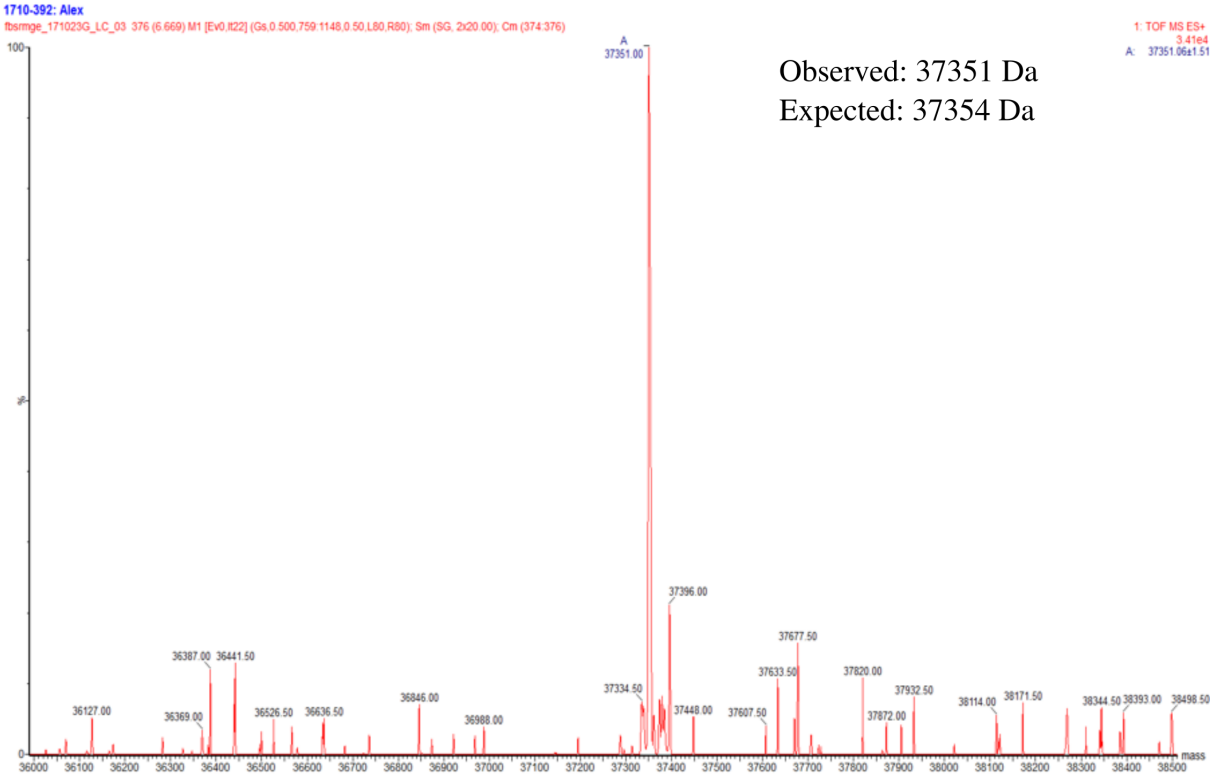


Figure A.22: Deconvoluted spectrum for Y30F R229N H83S F87Y eTA. The observed mass 37351 was within 3 Da of the expected mass for the variant (37354).

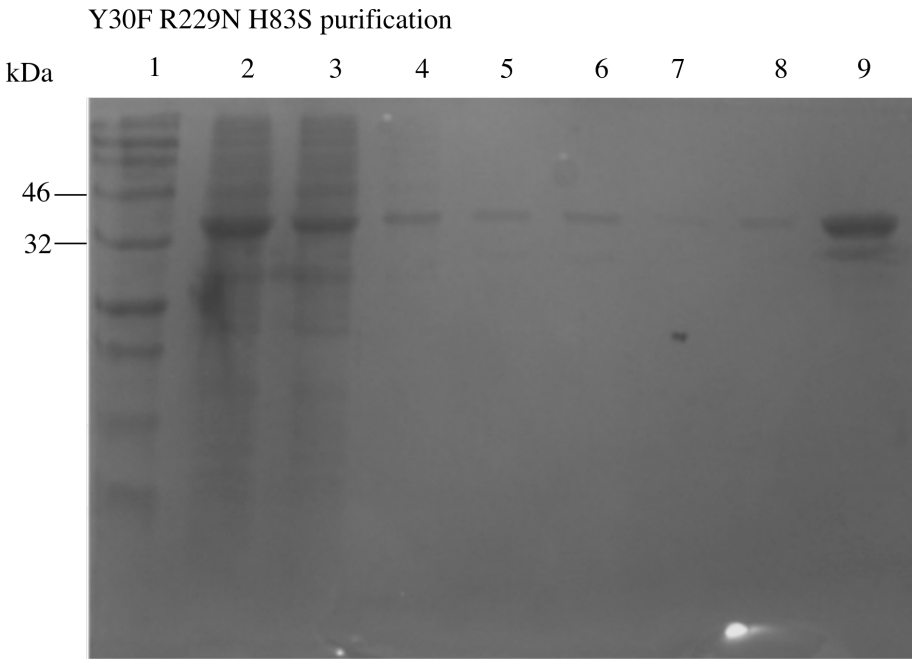


Figure A.23: SDS PAGE analysis of fractions from the purification of Y30F R229N H83S eTA. Lanes: 1 - protein ladder, 2 - soluble, 3 - insoluble, 4 - unbound, 5 - wash 1, 6 - wash 2, 7 - wash 3, 8 - wash 4, 9 - elution at 20 times dilute, 10 - elution at 40 times dilute.

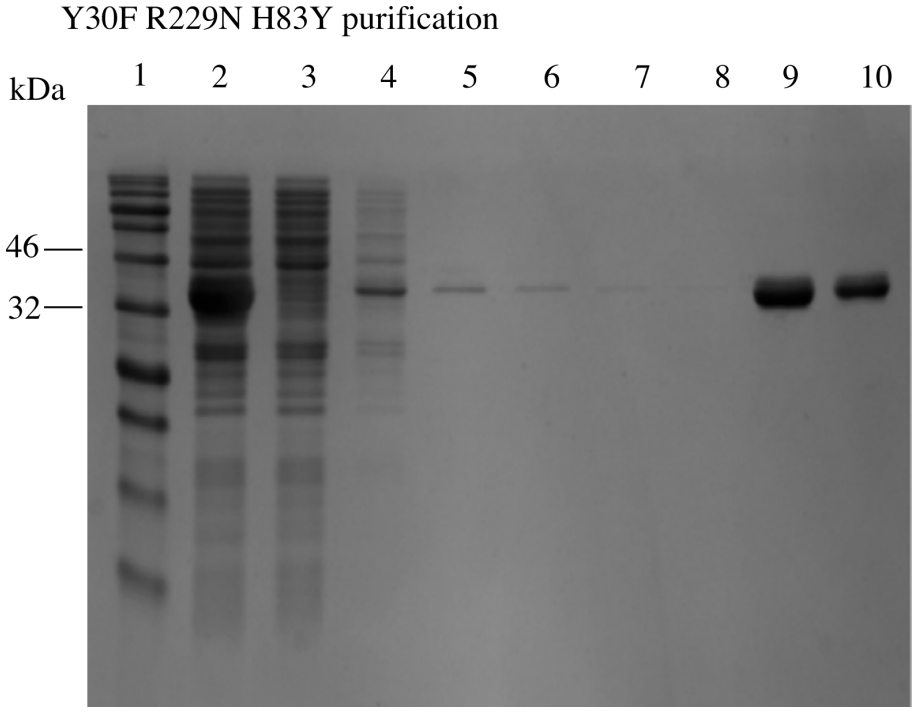


Figure A.24: SDS PAGE analysis of fractions from the purification of Y30F R229N F87Y eTA. Lanes: 1 - protein ladder, 2 - soluble, 3 - insoluble, 4 - unbound, 5 - wash 1, 6 - wash 2, 7 - wash 3, 8 - wash 4, 9 - elution at 20 times dilute, 10 - elution at 40 times dilute.

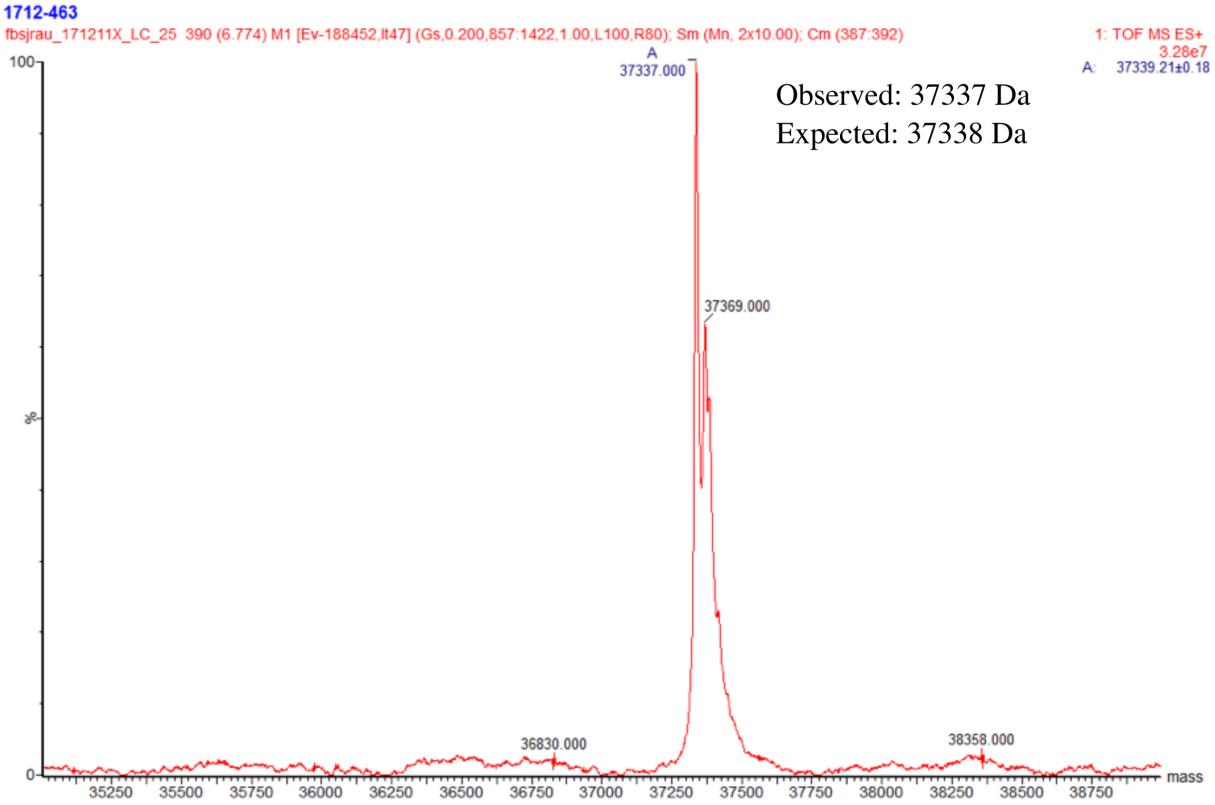


Figure A.25: Deconvoluted spectrum for Y30F R229N H83S eTA. The observed mass 37337 was within 1 Da of the expected mass for the variant (37338).

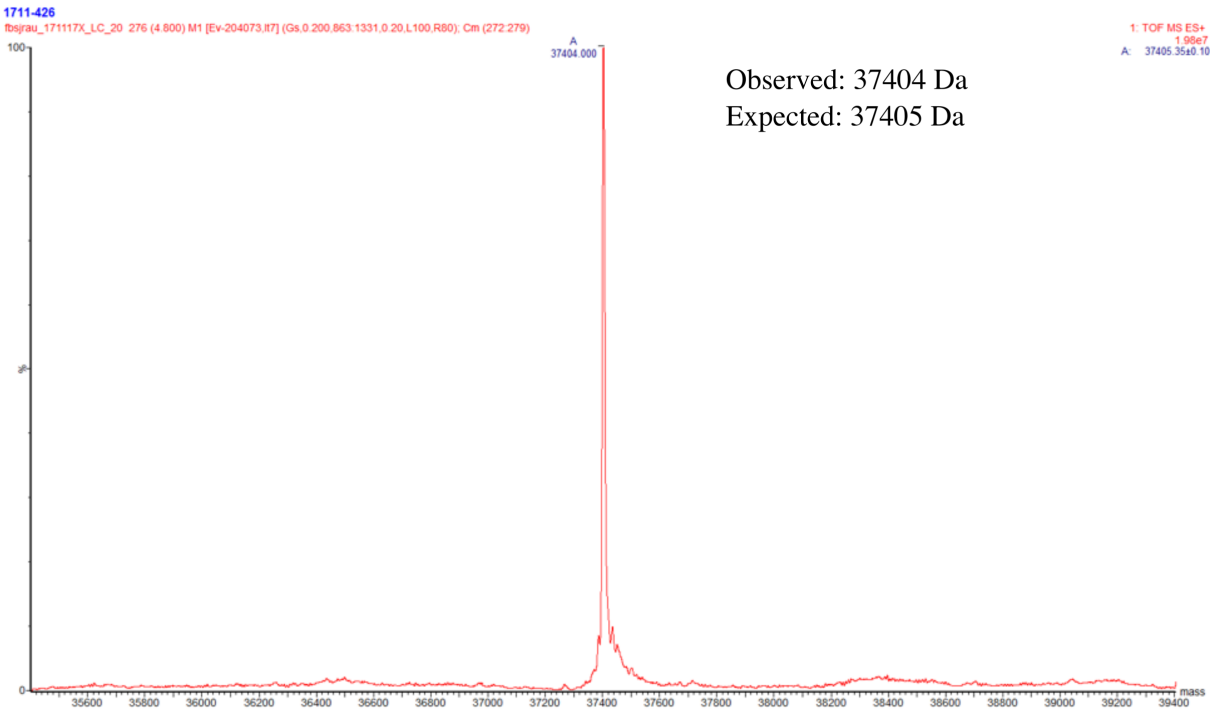


Figure A.26: Deconvoluted spectrum for Y30F R229N F87Y eTA. The observed mass 37404 was within 1 Da of the expected mass for the variant (37405).

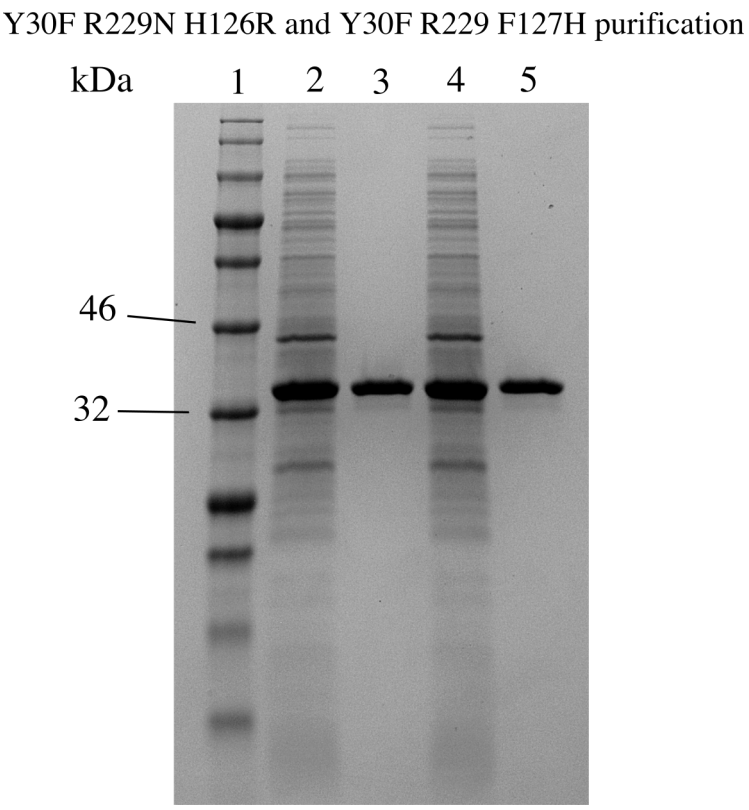


Figure A.27: SDS PAGE analysis for Y30F R229N H126R and Y30F R229N F127H eTA variants. Lanes: 1 - protein ladder, 2 - Y30F R229N H126R soluble, 3 - Y30F R229N H126R elution, 4 - Y30F R229N F127H soluble, 5 - Y30F R229N F127H elution.

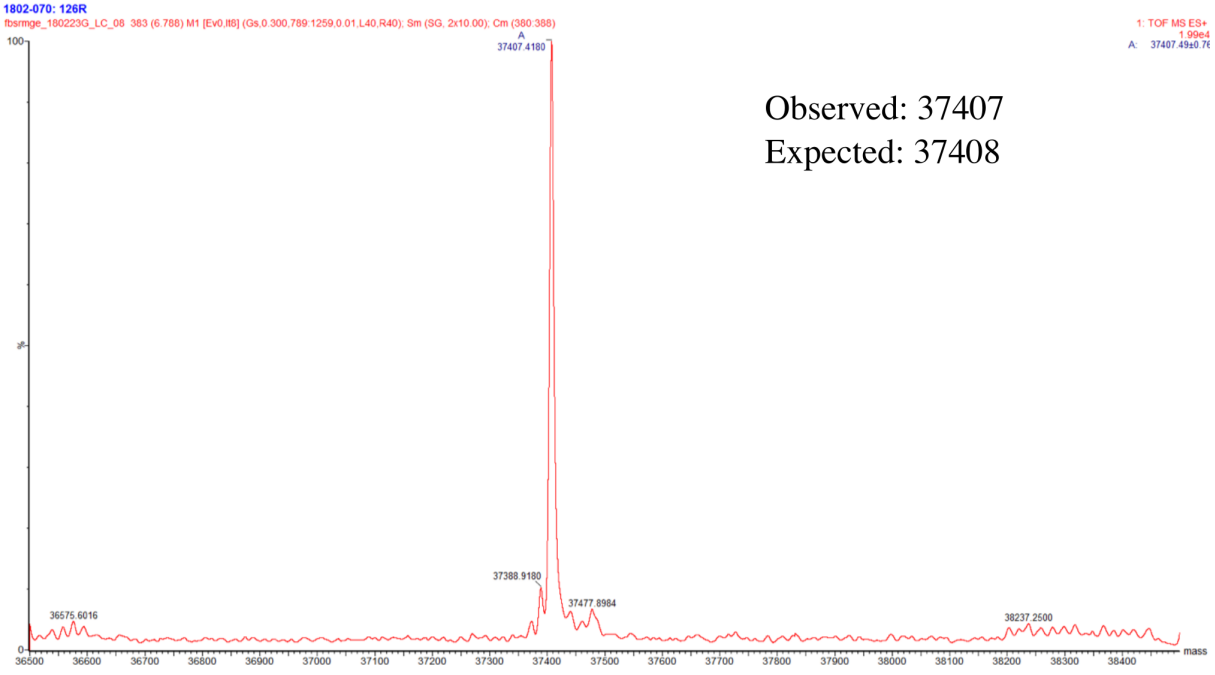


Figure A.28: Deconvoluted spectrum for Y30F R229N H126R eTA. The observed mass 37407 was within 1 Da of the expected mass for the variant (37408).

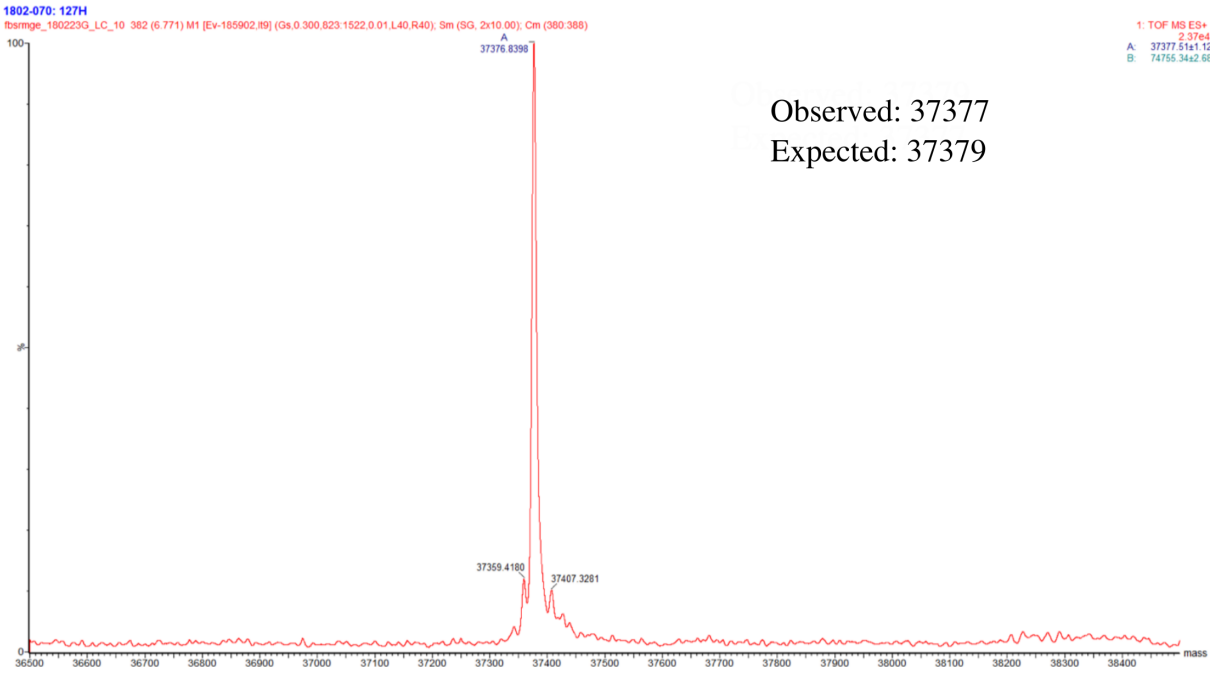


Figure A.29: Deconvoluted spectrum for Y30F R229N F127H eTA. The observed mass 37377 was within 2 Da of the expected mass for the variant (37379).

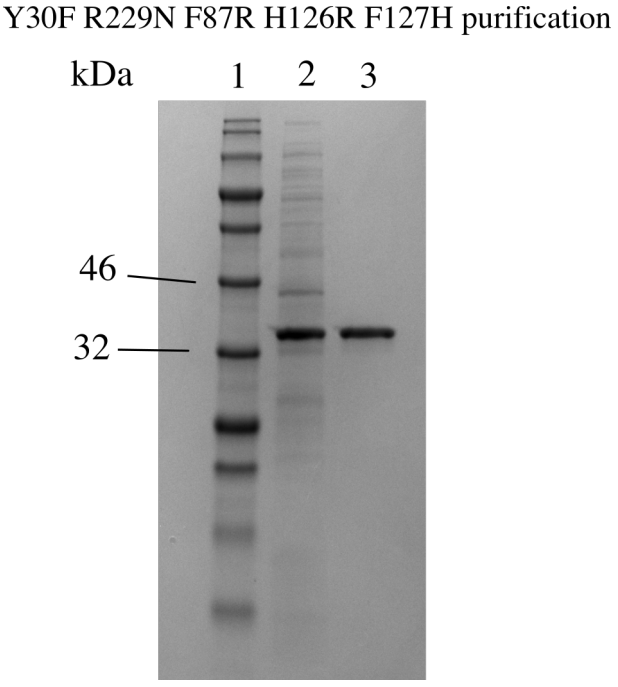


Figure A.30: SDS PAGE analysis for Y30F R229N F87R H126R F127H eTA. Lanes: 1 - protein ladder, 2 - soluble fraction, 3 - eluted fraction

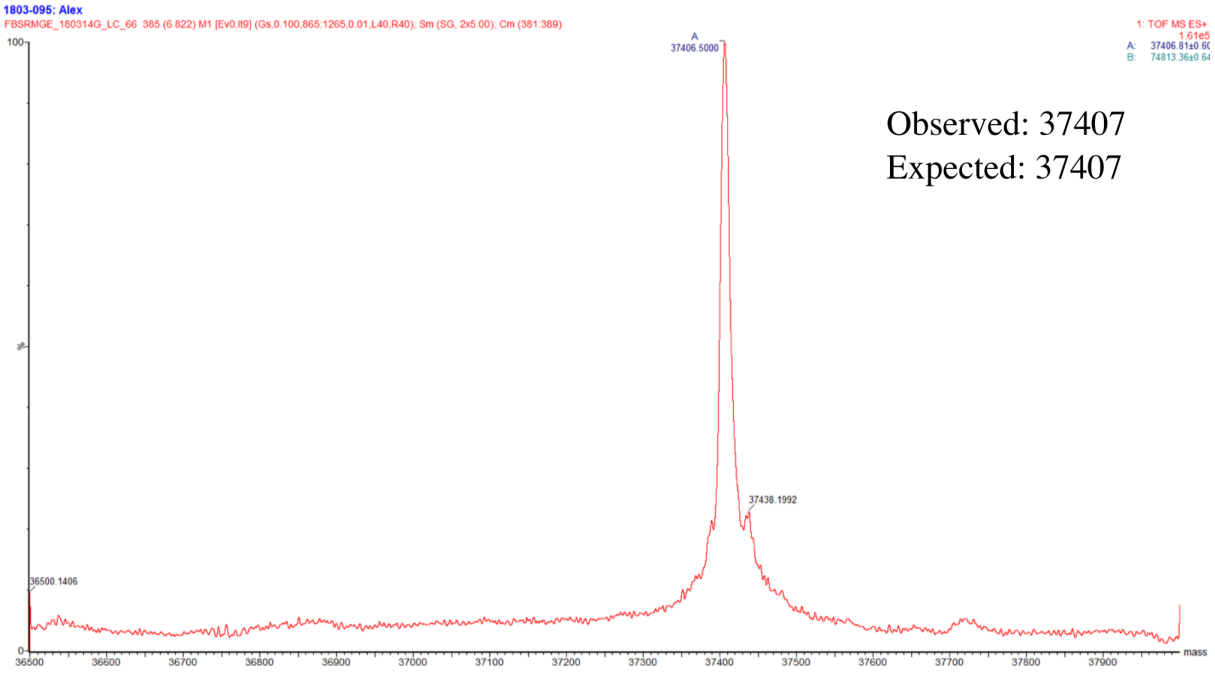


Figure A.31: Deconvoluted spectrum for Y30F R229N F87R H126R F127H eTA. The observed mass 37407 was within 1 Da of the expected mass for the variant (37408).

Bibliography

J. J. Agresti, E. Antipov, A. R. Abate, K. Ahn, A. C. Rowat, J.-C. Baret, M. Marquez, A. M. Klibanov, A. D. Griffiths, and D. A. Weitz. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proceedings of the National Academy of Sciences*, 2010.

E. A. Althoff, L. Wang, L. Jiang, L. Giger, J. K. Lassila, Z. Wang, M. Smith, S. Hari, P. Kast, D. Herschlag, et al. Robust design and optimization of retroaldol enzymes. *Protein Science*, 21:717–726, 2012.

S. F. Beaudoin, M. P. Hanna, I. Ghiviriga, and J. D. Stewart. Progress in using threonine aldolases for preparative synthesis. *Enzyme and Microbial Technology*, 119:1–9, 2018.

R. Blomberg, H. Kries, D. M. Pinkas, P. R. Mittl, M. G. Grütter, H. K. Privett, S. L. Mayo, and D. Hilvert. Precision is essential for efficient catalysis in an evolved kemp eliminase. *Nature*, 503:418, 2013.

U. T. Bornscheuer and R. J. Kazlauskas. *Hydrolases in organic synthesis: regio-and stereoselective biotransformations*. John Wiley & Sons, 2006.

U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, and K. Robins. Engineering the third wave of biocatalysis. *Nature*, 485:185–94, 2012.

A. L. Bowman, I. M. Grant, and A. J. Mulholland. QM/MM simulations predict a covalent intermediate in the hen egg white lysozyme reaction with its natural substrate. *Chemical Communications*, pages 4425–4427, 2008.

R. C. Cadwell and G. F. Joyce. Randomization of genes by PCR mutagenesis. *Genome Research*, 2:28–33, 1992.

K. Chen and F. H. Arnold. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proceedings of the National Academy of Sciences*, 90:5618–5622, 1993.

Q. Chen, X. Chen, Y. Cui, J. Ren, W. Lu, J. Feng, Q. Wu, and D. Zhu. A new D-threonine aldolase as a promising biocatalyst for highly stereoselective preparation of chiral aromatic β-hydroxy-α-amino acids. *Catalysis Science & Technology*, 7:5964–5973, 2017.

Y. Chen, S. L. Goldberg, R. L. Hanson, W. L. Parker, I. Gill, T. P. Tully, M. A. Montana, A. Goswami, and R. N. Patel. Enzymatic preparation of an (s)-amino acid from a racemic amino acid. *Organic Process Research & Development*, 15:241–248, 2010.

M. T. Christen, L. Menon, N. S. Myshakina, J. Ahn, M. A. Parniak, and R. Ishima. Structural basis of the allosteric inhibitor interaction on the HIV-1 reverse transcriptase RNase H domain. *Chemical Biology & Drug Design*, 80:706–716, 2012.

F. Claeyssens, K. E. Ranaghan, N. Lawan, S. J. Macrae, F. R. Manby, J. N. Harvey, and A. J. Mulholland. Analysis of chorismate mutase catalysis by QM/MM modelling of enzyme-catalysed and uncatalysed reactions. *Organic & Biomolecular Chemistry*, 9: 1578–1590, 2011.

P. Clapes, W. D. Fessner, G. A. Sprenger, and A. K. Samland. Recent progress in stereoselective synthesis with aldolases. *Current Opinions in Chemical Biology* , 14:154–67, 2010.

S. Dajnowicz, R. C. Johnston, J. M. Parks, M. P. Blakeley, D. A. Keen, K. L. Weiss, O. Gerlits, A. Kovalevsky, and T. C. Mueser. Direct visualization of critical hydrogen atoms in a pyridoxal 5'-phosphate enzyme. *Nature Communications*, 8:955, 2017.

A. D. Daniels, I. Campeotto, M. W. van der Kamp, A. H. Bolt, C. H. Trinh, S. E. Phillips, A. R. Pearson, A. Nelson, A. J. Mulholland, and A. Berry. Reaction mechanism of N-acetylneuraminic acid lyase revealed by a combination of crystallography, QM/MM simulation, and mutagenesis. *ACS Chemical Biology*, 9:1025–1032, 2014.

I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall III, J. Snoeyink, J. S. Richardson, et al. Molprobity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*, 35:375–383, 2007.

H. A. De Boer, L. J. Comstock, and M. Vasser. The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proceedings of the National Academy of Sciences*, 80:21–25, 1983.

W. DeLano. The pymol molecular graphics system; delano scientific: San carlos, ca, 2002. *PyMOL*, 2009.

M. L. di Salvo, S. G. Remesh, M. Vivoli, M. S. Ghatge, A. Paiardini, S. D'Aguanno, M. K. Safo, and R. Contestabile. On the catalytic mechanism and stereospecificity of *Escherichia coli* L-threonine aldolase. *The Federation of European Biochemical Societies Journal*, 281: 129–45, 2014.

W. Diniz and F. Canduri. Bioinformatics: an overview and its applications. *Genetics and Molecular Research*, 16, 2017.

P. Emsley and K. Cowtan. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, 60:2126–32, 2004.

P. Emsley, B. Lohkamp, W. G. Scott, and K. Cowtan. Features and development of coot. *Acta Crystallographica Section D: Biological Crystallography*, 66:486–501, 2010.

D. A. Estell, T. P. Graycar, and J. A. Wells. Engineering an enzyme by site-directed mutagenesis to be resistant to chemical oxidation. *Journal of Biological Chemistry*, 260: 6518–6521, 1985.

K. Fesko, D. Suplatov, and V. Švedas. Bioinformatic analysis of the fold type I PLP-dependent enzymes reveals determinants of reaction specificity in L-threonine aldolase from *Aeromonas jandaei*. *Federation of European Biochemical Societies open bio*, 8:1013–1028, 2018.

W.-D. Fessner, A. Schneider, H. Held, G. Sinerius, C. Walter, M. Hixon, and J. V. Schloss. The mechanism of class II, metal-dependent aldolases. *Angewandte Chemie International Edition in English*, 35:2219–2221, 1996.

S. E. Franz and J. D. Stewart. *Chapter Three - Threonine Aldolases*, volume 88, pages 57–101. Academic Press, 2014.

T. Gefflaut, C. Blonski, J. Perie, and M. Willson. Class I aldolases: substrate specificity, mechanism, inhibitors and structural aspects. *Progress in Biophysics and Molecular Biology*, 63:301–340, 1995.

L. Giger, M. D. Toscano, M. Bouzon, P. Marlière, and D. Hilvert. A novel genetic selection system for PLP-dependent threonine aldolases. *Tetrahedron*, 68:7549–7557, 2012.

L. Giger, S. Caner, R. Obexer, P. Kast, D. Baker, N. Ban, and D. Hilvert. Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nature Chemical Biology*, 9:494, 2013.

H. Griengl, H. Schwab, and M. Fechter. The synthesis of chiral cyanohydrins by oxynitrilases. *Trends in Biotechnology*, 18:252–256, 2000.

M. T. Guo, A. Rotem, J. A. Heyman, and D. A. Weitz. Droplet microfluidics for high-throughput biological assays. *Lab on a Chip*, 12:2146–2155, 2012.

H. J. Gwon and S. H. Baik. Diastereoselective synthesis of L-*threo*-3,4-dihydroxyphenylserine by *low-specificity* L-threonine aldolase mutants. *Biotechnology Letters*, 32:143–9, 2010.

H. J. Gwon, H. Yoshioka, N. E. Song, J. H. Kim, Y. R. Song, D. Y. Jeong, and S. H. Baik. Optimal production of L-*threo*-3,4-dihydroxyphenylserine (L-*threo*-dops) on a large scale by diastereoselectivity-enhanced variant of L-threonine aldolase expressed in *Escherichia coli*. *Preparative Biochemistry and Biotechnology*, 42:143–54, 2012.

K. B. Hansen, Y. Hsiao, F. Xu, N. Rivera, A. Clausen, M. Kubryk, S. Krska, T. Rosner, B. Simmons, J. Balsells, et al. Highly efficient asymmetric synthesis of sitagliptin. *Journal of the American Chemical Society*, 131:8798–8804, 2009.

H. Hayashi, H. Mizuguchi, and H. Kagamiyama. The imine- pyridine torsion of the pyridoxal 5 '-phosphate Schiff base of aspartate aminotransferase lowers its pk_a in the unliganded enzyme and is crucial for the successive increase in the pk_a during catalysis. *Biochemistry*, 37:15076–15085, 1998.

H. Hayashi, H. Mizuguchi, I. Miyahara, Y. Nakajima, K. Hirotsu, and H. Kagamiyama. Conformational change in aspartate aminotransferase on substrate binding induces strain

in the catalytic group and enhances catalysis. *Journal of Biological Chemistry*, 278: 9481–9488, 2003.

R. B. Herbert, B. Wilkinson, and G. J. Ellames. Preparation of (2R, 3S)-β-hydroxy-α-amino acids by use of a novel *Streptomyces* aldolase as a resolving agent for racemic material. *Canadian Journal of Chemistry*, 72:114–117, 1994.

M. Höhne, S. Schätzle, H. Jochens, K. Robins, and U. T. Bornscheuer. Rational assignment of key motifs for function guides *in silico* enzyme identification. *Nature chemical biology*, 6:807, 2010.

S. Jennewein, M. Schürmann, M. Wolberg, I. Hilker, R. Luiten, M. Wubbolts, and D. Mink. Directed evolution of an industrial biocatalyst: 2-deoxy-D-ribose 5-phosphate aldolase. *Biotechnology Journal: Healthcare Nutrition Technology*, 1:537–548, 2006.

L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas, et al. *De novo* computational design of retro-aldol enzymes. *Science*, 319:1387–1391, 2008.

H. Jochens and U. T. Bornscheuer. Natural diversity to guide focused directed evolution. *ChemBioChem*, 11:1861–1866, 2010.

S. C. Kamerlin and A. Warshel. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins: Structure, Function, and Bioinformatics*, 78:1339–1375, 2010.

O. Khersonsky, D. Röthlisberger, O. Dym, S. Albeck, C. J. Jackson, D. Baker, and D. S. Tawfik. Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the ke07 series. *Journal of Molecular Biology*, 396:1025–1042, 2010.

C. L. Kielkopf and S. K. Burley. X-ray structures of threonine aldolase complexes: structural basis of substrate recognition. *Biochemistry*, 41:11711–20, 2002.

S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z.-G. Zhang, D. J. Opperman, M. T. Reetz, and J. P. Acevedo. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synthetic Biology*, 2:83–92, 2012.

J. Kim, J. P. Kershner, Y. Novikov, R. K. Shoemaker, and S. D. Copley. Three serendipitous pathways in *E. coli* can bypass a block in pyridoxal-5'-phosphate synthesis. *Molecular Systems Biology*, 6:88, 2010.

J. R. Knowles. Enzyme catalysis: not different, just better. *Nature*, 350:121, 1991.

H. Kries, R. Blomberg, and D. Hilvert. *De novo* enzymes by computational design. *Current Opinion in Chemical Biology*, 17:221–228, 2013.

B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, and D. Baker. Design of a novel globular protein fold with atomic-level accuracy. *Science*, 302:1364–1368, 2003.

S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, 13:1011–1021, 1992.

D. W. Leung. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique*, 1:11–15, 1989.

C. Li, A. Wen, B. Shen, J. Lu, Y. Huang, and Y. Chang. Fastcloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. *BMC Biotechnology*, 11:92, 2011.

J. Liu, M. Odani, T. Yasuoka, T. Dairi, N. Itoh, M. Kataoka, S. Shimizu, and H. Yamada. Gene cloning and overproduction of low-specificity L-*allo* aldolase from alcaligenes xylosoxidans and its application for production of a key intermediate for parkinsonism drug. *Applied Microbiology and Biotechnology*, 54:44–51, 2000.

J. Q. Liu, T. Dairi, M. Kataoka, S. Shimizu, and H. Yamada. L-*allo*-threonine aldolase from *Aeromonas jandaei* dk-39: gene cloning, nucleotide sequencing, and identification of the pyridoxal 5'-phosphate-binding lysine residue by site-directed mutagenesis. *J Bacteriol*, 179:3555–60, 1997.

J. Q. Liu, T. Dairi, N. Itoh, M. Kataoka, S. Shimizu, and H. Yamada. Gene cloning, biochemical characterization and physiological role of a thermostable low-specificity L-threonine aldolase from *Escherichia coli*. *European Journal of Biochemistry*, 255:220–6, 1998.

A. Lodola, M. Mor, J. Zurek, G. Tarzia, D. Piomelli, J. N. Harvey, and A. J. Mulholland. Conformational effects in enzyme catalysis: reaction via a high energy conformation in fatty acid amide hydrolase. *Biophysical Journal*, 92:L20–L22, 2007.

F. Ma, M. T. Chung, Y. Yao, R. Nidetz, L. M. Lee, A. P. Liu, Y. Feng, K. Kurabayashi, and G.-Y. Yang. Efficient molecular evolution to generate enantioselective enzymes using a dual-channel microfluidic droplet screening platform. *Nature communications*, 9:1030, 2018.

T. D. Machajewski and C. H. Wong. The catalytic asymmetric aldol reaction. *Angewandte Chemie International Edition English*, 39:1352–1375, 2000.

J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *Journal of Chemical Theory and Computation*, 11:3696–3713, 2015.

H. Matsumae, M. Furui, and T. Shibatani. Lipase-catalyzed asymmetric hydrolysis of 3-phenylglycidic acid ester, the key intermediate in the synthesis of diltiazem hydrochloride. *Journal of Fermentation and Bioengineering*, 75:93–98, 1993.

A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. Phaser crystallographic software. *Journal of applied crystallography*, 40:658–674, 2007.

I. Moreno-Gonzalez and C. Soto. Misfolded protein aggregates: mechanisms, structures and potential for disease transmission. In *Seminars in cell & developmental biology*, volume 22, pages 482–487. Elsevier, 2011.

P. F. Mugford, U. G. Wagner, Y. Jiang, K. Faber, and R. J. Kazlauskas. Enantiocomplementary enzymes: Classification, molecular basis for their enantiopreference, and prospects for mirror-image biotransformations. *Angewandte Chemie International Edition*, 47: 8782–8793, 2008.

M. Müller. Chemoenzymatic synthesis of building blocks for statin side chains. *Angewandte Chemie International Edition*, 44:362–365, 2005.

G. N. Murshudov, A. A. Vagin, and E. J. Dodson. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D*, 53:240–55, 1997.

R. Obexer, A. Godina, X. Garrabou, P. R. Mittl, D. Baker, A. D. Griffiths, and D. Hilvert. Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nature Chemistry*, 9:50, 2017.

N. Ogawa, M. Yamamoto, and H. Takayama. L-*threo*-3, 4-dihydroxyphenylserine treatment of Parkinson's disease. *Journal of medicine*, 16:525–534, 1985.

A. Paiardini, R. Contestabile, S. D'Aguanno, S. Pascarella, and F. Bossa. Threonine aldolase and alanine racemase: novel examples of convergent evolution in the superfamily of vitamin B6-dependent enzymes. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1647:214–219, 2003.

T. Palzkill and D. Botstein. Identification of amino acid substitutions that alter the substrate specificity of TEM-1 beta-lactamase. *Journal of Bacteriology*, 174:5237–5243, 1992.

R. N. Patel. Biocatalytic synthesis of chiral alcohols and amino acids for development of pharmaceuticals. *Biomolecules*, 3:741–777, 2013.

W. M. Patrick, A. E. Firth, and J. M. Blackburn. User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein engineering*, 16:451–457, 2003.

S. Prasad, M. Bocola, and M. T. Reetz. Revisiting the lipase from *Pseudomonas aeruginosa:* directed evolution of substrate acceptance and enantioselectivity using iterative saturation mutagenesis. *ChemPhysChem*, 12:1550–7, 2011.

H. M. Qin, F. L. Imai, T. Miyakawa, M. Kataoka, N. Kitamura, N. Urano, K. Mori, H. Kawabata, M. Okai, J. Ohtsuka, F. Hou, K. Nagata, S. Shimizu, and M. Tanokura. L-*allo*-threonine aldolase with an H128Y/S292R mutation from *Aeromonas jandaei* DK-39 reveals the structural basis of changes in substrate stereoselectivity. *Acta Crystallographica Section D*, 70:1695–703, 2014.

R. J. Ratzan, M. A. Moore, and A. A. Yunis. Effect of chloramphenicol and thiamphenicol on the *in vitro* colony-forming cell. *Blood*, 43:363–369, 1974.

M. T. Reetz. Controlling the enantioselectivity of enzymes by directed evolution: practical and theoretical ramifications. *Proceedings of the National Academy of Sciences of the United States of America*, 101:5716–22, 2004.

M. T. Reetz, S. Wilensek, D. Zha, and K.-E. Jaeger. Directed evolution of an enantioselective enzyme through combinatorial multiple-cassette mutagenesis. *Angewandte Chemie International Edition*, 40:3589–3591, 2001.

M. T. Reetz, C. Torre, A. Eipper, R. Lohmer, M. Hermes, B. Brunner, A. Maichele, M. Bocola, M. Arand, A. Cronin, et al. Enhancing the enantioselectivity of an epoxide hydrolase by directed evolution. *Organic letters*, 6:177–180, 2004.

M. T. Reetz, M. Bocola, J. D. Carballeira, D. Zha, and A. Vogel. Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angewandte Chemie*, 117:4264–4268, 2005.

M. T. Reetz, L.-W. Wang, and M. Bocola. Directed evolution of enantioselective enzymes: Iterative cycles of CASTing for probing protein-sequence space. *Angewandte Chemie*, 118: 1258–1263, 2006.

M. T. Reetz, D. Kahakeaw, and R. Lohmer. Addressing the numbers problem in directed evolution. *ChemBioChem*, 9:1797–1804, 2008.

C. Reisinger, F. van Assema, M. Schürmann, Z. Hussain, P. Remler, and H. Schwab. A versatile colony assay based on NADH fluorescence. *Journal of Molecular Catalysis B: Enzymatic*, 39:149 – 155, 2006.

C. Reuben, A. Marina, E. Alexis, E. T. S. C., D. M. J., and T. N. J. Directed evolution of an amine oxidase possessing both broad substrate specificity and high enantioselectivity. *Angewandte Chemie*, 115:4955–4958, 2008.

D. R. Roe and T. E. Cheatham III. Ptraj and cpptraj: software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, 9: 3084–3095, 2013.

L. Rosenthaler. Durch enzyme bewirkte asymmetrische synthesen. *Biochem. Z*, 14:238–253, 1908.

D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, et al. Kemp elimination catalysts by computational enzyme design. *Nature*, 453:190, 2008.

I. Rowles, B. Groenendaal, B. Binay, K. J. Malone, S. C. Willies, and N. J. Turner. Engineering of phenylalanine ammonia lyase from *Rhodotorula graminis* for the enhanced synthesis of unnatural L-amino acids. *Tetrahedron*, 72:7343–7347, 2016.

A. G. Sandström, Y. Wikmark, K. Engström, J. Nyhlén, and J.-E. Bäckvall. Combinatorial reshaping of the *Candida antarctica* lipase A substrate pocket for enantioselectivity using an extremely condensed library. *Proceedings of the National Academy of Sciences*, 109: 78–83, 2012.

C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, et al. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science*, 329:305–309, 2010.

Y. V. G. Schlippe and L. Hedstrom. A twisted base? The role of arginine in enzyme-catalyzed proton abstractions. *Archives of Biochemistry and Biophysics*, 433:266–278, 2005.

A. W. Schuttelkopf and D. M. van Aalten. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica Section D*, 60: 1355–63, 2004.

J. Shao, S. W. Tanner, N. Thompson, and T. E. Cheatham. Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *Journal of Chemical Theory and Computation*, 3:2312–2334, 2007.

P. E. Siegbahn and F. Himo. Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. *Journal of Biological Inorganic Chemistry*, 14:643–651, 2009.

J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. S. Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, et al. Computational design of an enzyme catalyst for a stereoselective bimolecular diels-alder reaction. *Science*, 329:309–313, 2010.

C. R. Søndergaard, M. H. Olsson, M. Rostkowski, and J. H. Jensen. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pk_a values. *Journal of Chemical Theory and Computation*, 7:2284–2295, 2011.

J. Steinreiber, K. Fesko, C. Reisinger, M. Schürmann, F. van Assema, M. Wolberg, D. Mink, and H. Griengl. Threonine aldolases—an emerging tool for organic synthesis. *Tetrahedron*, 63:918–926, 2007.

W. P. Stemmer. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proceedings of the National Academy of Sciences*, 91:10747–10751, 1994a.

W. P. Stemmer. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature*, 370:389, 1994b.

J. J. Stewart. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling*, 13:1173–1213, 2007.

J. D. Tibhe, H. Fu, T. Noël, Q. Wang, J. Meuldijk, and V. Hessel. Flow synthesis of phenylserine using threonine aldolase immobilized on Eupergit support. *Beilstein Journal of Organic Chemistry*, 9:2168, 2013.

G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23:187–199, 1977.

M. D. Truppo. Biocatalysis in the pharmaceutical industry: the need for speed. *ACS Medicinal Chemistry Letters*, 8:476–480, 2017.

N. J. Turner and M. D. Truppo. Biocatalysis enters a new era. *Current Opinion in Chemical Biology*, 2:212–214, 2013.

M. K. Uhl, G. Oberdorfer, G. Steinkellner, L. Riegler-Berket, D. Mink, F. van Assema, M. Schurmann, and K. Gruber. The crystal structure of D-threonine aldolase from *Alcaligenes xylosoxidans* provides insight into a metal ion assisted PLP-dependent mechanism. *PLoS One*, 10, 2015.

N. Urano, S. Fukui, S. Kumashiro, T. Ishige, S. Kita, K. Sakamoto, M. Kataoka, and S. Shimizu. Directed evolution of an aminoalcohol dehydrogenase for efficient production of double chiral aminoalcohols. *Journal of Bioscience and Bioengineering*, 111:266–271, 2011.

M. W. van der Kamp and A. J. Mulholland. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry*, 52: 2708–2728, 2013.

M. W. van der Kamp, F. Perruccio, and A. J. Mulholland. High-level QM/MM modelling predicts an arginine as the acid in the condensation reaction catalysed by citrate synthase. *Chemical Communications*, pages 1874–1876, 2008.

M. W. van der Kamp, J. D. McGeagh, and A. J. Mulholland. "Lethal Synthesis" of Fluorocitrate by Citrate Synthase Explained through QM/MM Modeling. *Angewandte Chemie*, 123:10533–10535, 2011.

A. Warshel. Computer simulations of enzyme catalysis: methods, progress, and insights. *Annual Review of Biophysics and Biomolecular Structure*, 32:425–443, 2003.

J. A. Wells, D. B. Powers, R. R. Bott, T. P. Graycar, and D. A. Estell. Designing substrate specificity by protein engineering of electrostatic interactions. *Proceedings of the National Academy of Sciences*, 84:1219–1223, 1987.

G. J. Williams, S. Domann, A. Nelson, and A. Berry. Modifying the stereochemistry of an enzyme-catalyzed reaction by directed evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 100:3143–8, 2003.

D. Williams-Herman, E. Round, A. S. Swern, B. Musser, M. J. Davies, P. P. Stein, K. D. Kaufman, and J. M. Amatruda. Safety and tolerability of sitagliptin in patients with type 2 diabetes: a pooled analysis. *BMC Endocrine Disorders*, 8:14, 2008.

D. S. Wilson and A. D. Keefe. Random mutagenesis by PCR. *Current Protocols in Molecular Biology*, 51:8–3, 2000.

C. L. Windle, M. Muller, A. Nelson, and A. Berry. Engineering aldolases as biocatalysts. *Current Opinion in Chemical Biology*, 19:25–33, 2014.

G. Winter. Xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography*, 43:186–190, 2010.

C. H. Wong and G. M. Whitesides. Synthesis of sugars by aldolase-catalyzed condensation reactions. *The Journal of Organic Chemistry*, 48:3199–3205, 1983.

G. Yang and S. G. Withers. Ultrahigh-throughput FACS-based screening for directed enzyme evolution. *ChemBioChem*, 10:2704–2715, 2009.

H. Zhao, L. Giver, Z. Shao, J. A. Affholter, and F. H. Arnold. Molecular evolution by staggered extension process (StEP) *in vitro* recombination. *Nature Biotechnology*, 16:258, 1998.